Updates in Metabolomics Tools and Resources: 2014-2015

Journal:	*ELECTROPHORESIS*
Manuscript ID	elps.201500417.R1
Wiley - Manuscript type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Misra, Biswapriya; University of Florida, Biology; van der Hooft, Justin; University of Glasgow, College of Medical, Veterinary and Life Sciences;
Keywords:	annotation, databases, mass spectrometry, metabolites, metabolomics

Updates in Metabolomics Tools and Resources: 2014-2015

Biswapriya B. Misra[1,*] and Justin J.J. van der Hooft[2]

[1] Department of Biology, Genetics Institute, University of Florida, Gainesville, FL 32610, USA

[2]Glasgow Polyomics, University of Glasgow, Glasgow, G61 1QH, UK

***Corresponding author**:

Biswapriya B. Misra, Ph.D.

Cancer & Genetics Research Complex, Room 437

University of Florida

2033 Mowry Road, Gainesville

FL 32610, USA

Tel: +1 (352) 215-6040

E-mail: biswapriyamisra@ufl.edu

Abbreviations:

HMDB: Human Metabolome Database; KEGG: Kyoto Encyclopedia of Genes and Genomes; MRM: multiple reaction monitoring; PCA: principal component analysis; PLS-DA: partial least square; PRIMe: Platform for RIKEN Metabolomics;

Keywords: annotation, databases, mass spectrometry, metabolites, metabolomics, NMR, statistics, software tools, data processing, data analysis, data visualization.

Total Number of words: 12, 764

Abstract

Data processing and interpretation represent the most challenging and time-consuming steps in high-throughput metabolomic experiments, regardless of the analytical platform (mass spectrometry [MS] or nuclear magnetic resonance spectroscopy [NMR]-based) used for data acquisition. Improved machinery in metabolomics generate increasingly complex data sets which create the need for more and better processing and analysis software and *in-silico* approaches to understand the resulting data. However, a comprehensive source of information describing the utility of the most recently developed and released metabolomics resources -- in the form of tools, software, and databases - is currently lacking. Thus, here we provide an overview of freely-available, open-source, tools, algorithms and frameworks to make both upcoming and established metabolomics researchers aware of the recent developments in an attempt to advance and facilitate data processing workflows in their metabolomics research. The major topics include tools and researches for data processing, data annotation, and data visualization in MS and NMR based metabolomics. Most in this review described tools are dedicated to untargeted metabolomics workflows; however, some more specialist tools are described as well. All tools and resources described including their analytical and computational platform dependencies are summarized in an overview Table.

1. Introduction

In metabolomics, data processing and interpretation represent some of the most challenging and time-consuming steps in the high throughput process, regardless of the analytical platform used for data generation. The exponentially growing volume of generated data has triggered the research and development of tools, software, programs, databases, and applications to facilitate the robust understanding of metabolic processes of biological systems. For instance, natural products discovery has greatly benefited from the development of open-access spectral and chemical databases [1]. In addition, a large number of chemoinformatics tools are finding direct application in handling metabolomics data or helping in annotation. Targeted metabolomics investigations obtain quantitative data on a predefined set of compounds, while untargeted metabolomics studies provide a broader exploration of metabolites with the goal of identifying new compounds [2]. Untargeted metabolomic studies are characterized by simultaneous qualitative and quantitative analysis of a large number of metabolites in samples. Currently, untargeted metabolomics is being increasingly applied in diverse areas of research such as biotechnology, understanding of microbial interactions, human health and disease, and functional genomics in the form of metabolic genome-wide associated studies (mGWAS); however, all these applications share the same bottlenecks: i.e., the harmonization and coverage of existing analytical methods, the lack of automation of spectral data processing, and the data interpretation [3]. These limitations are the major driving force providing impetus for the development of a huge plethora of metabolomics tools, software, programs, and databases. Although excellent compilations are available for MS-based proteomics tools and frameworks [4], comparable efforts in metabolomics are few [5]. Online resources such as OMICtools (http://omictools.com/) [6], Fiehn Lab resources [7], Metabolomics Society's resource pages [8] and meeting highlights [9], and a recently started metabomatch [10] catalog and archive are some of the major available resources. However, there are tools that were developed and released in the last two years that have not yet been described and listed. Given the tremendous challenge to accomplish multiple functions in limited time and in a more efficient manner using a single tool or approach, most of these tools perform multiple analyses steps: these new tools take care of basic but complex analysis functions and allow end-users to focus on the more sophisticated functions, data-specific analysis, and novel aspects of the data. Thus, the groupings of metabolomics tools in the following sections should serve more as guidance for readers than as

concrete classifications. An overview, focusing on the most recently published, of these freely available, open-source, downloadable tools, algorithms and frameworks is provided to make the community aware of them and in an attempt to advance and facilitate data processing workflow in metabolomics. Since the majority of tools were found to be developed for mass spectrometry data, this forms the large part of the review, with separate sub-sections devoted to GC-MS and NMR based tools. All tools and resources, including their analytical and computational platform dependencies, are summarized in Table 1.

2. Data handling and preprocessing

The two most commonly used approaches for the generation of metabolomic data are nuclear magnetic resonance (NMR) and mass spectrometry (MS) [11]. MS-based technologies used for high-throughput metabolomics include chromatography-based MS techniques, such as GC-MS and LC-MS, as well as chromatography-free methods, i.e., direct infusion, matrix-assisted, and matrix-free laser desorption/ionization, imaging, and other ambient ionization approaches among others. Metabolomics data handling and processing tools are critical for generating meaningful scientific results from the vast amount of raw data generally acquired in a typical metabolomic analysis. Data processing is the first step post-acquisition of raw data from the instruments/ platforms before starting any metabolite annotation or statistical analysis by facilitating an easy access to the characteristics of each observed feature, i.e., retention time, m/z values, ion intensity, and isotope distribution in original raw data sets. Mass spectrum pre-processing algorithms belong to five types of popular methods: spectrum normalization, spectrum clustering, precursor charge determination, spectrum de-noising, and spectrum quality assessment [12]. Most preprocessing software pipelines share the general functions of peak detection, peak alignment, and peak annotation, though the capabilities, advantages, and limitations vary dramatically among currently available software. For instance, innovations in peak picking, using wavelet-based peak picker (CantWaiT) and a precursor charge determination algorithm (Turbocharger), into ProteoWizard suite provide universal tools that are vendor-neutral for MS/MS data handling since they are based on the general mzML and mzXML data formats [13]. While there are established and time-tested tools that are already widely used for metabolomic data analysis, there is a wealth of software packages which offer specialized

functions and capabilities for researchers. There are multiple, popular open-source programs for MS peak picking, alignment, and annotation such as MZmine [14] which require manual inputs from the user at each step, in both file-to-file basis and in batch modes for large number of files. In batch-mode, the chosen options for defined, step-wise data analysis settings can be saved for future reproducible analysis. Alternatively, XCMS [15] requires more programming knowledge, but the data analysis steps can be automated once the appropriate parameters are chosen; however, recently, a web version (XCMS Online) was made available as well, which needs no programming knowledge to use. In addition, TracMass 2 is a suite of computer programs for full-scan MS data which provides graphical feedback to the data analyst on parameter settings and processes results in the form of a table of peak intensities [16]. This modular tool holds promise for handling of raw LC-MS data for both metabolomics as well as proteomics applications. MetMSLine, implemented in R helps achieve improved standard operating procedures and good laboratory practices for efficient and objective, untargeted metabolomics analyses of large-scale (e.g. >300 samples), high-resolution LC-MS datasets [17]. This R-based tool is compatible with any peak picking software output (e.g. XCMS, MZmine, and proprietary software) and is available as both a user friendly graphical user interface for non R specialists and in the form of command line functions for seamless integration. A new peak detection approach was implemented as part of the apLCMS package that learns directly from data features of extracted ion chromatograms (EICs) to differentiate true peaks from noise in LC-MS profiles [18].

In an interesting comparative study, four freely available and published preprocessing tools- MetAlign, MZmine, SpectConnect, and XCMS, were evaluated for impurity profiling using nominal mass GC/MS data and accurate mass LC/MS data. The study indicated that for GC/MS data, MetAlign had the most component detections, followed by MZmine, SpectConnect, and XCMS tools [19]. For the LC-MS data, the order was MetAlign, XCMS, and MZmine, whereas greater detection percentages (increased by ~ 15%) were obtained by combining the top performer with at least one of the other tools.

2.1 Tools for targeted analysis by multiple reaction monitoring

Large-scale, targeted metabolomics methods require different software functions and do generally not suffer from the same challenges as untargeted experiments; however, basic data

processing steps are needed to collect the relevant information from the spectral data. Recently, some tools dedicated to targeted metabolomics experiments were released. MRM-DIFF is software for differential analysis of large-scale multiple reaction monitoring (MRM) assays that uses a correlation-optimized warping (COW) algorithm to align MRM chromatograms, with support from quality control (QC) sample datasets to automatically adjust the alignment parameters [20]. Further, MRM-DIFF allows user-customized reference libraries, which can include molecular formula, retention time, and MRM transition information, for identification of the target molecules. A related software tool, MRMPROBS, aids in the analysis of large-scale MRM assays and supports the data formats of most major mass spectrometer vendors. MRMPROBS uses mzXML data format and provides a process pipeline from the raw data import to statistical analysis. Like other pipelines, MRMPROBS performs peak quantification, missing values imputation, and peak normalization [21]. Similarly, MRM-Ion Pair Finder software acquires characteristic MRM ion pairs by precursor ion alignment, MS2 spectrum extraction and reduction, characteristic product ion selection, and ion fusion [22]. 'FragPred' provided an *in silico* framework for designing QqQ MRM experiments for each of the 82,696 metabolites available in the METLIN metabolite database to enable the application of QqQ mass spectrometers to large-scale metabolite profiling using MRM experiments [23]. Another interesting recent tool, Multi-platform Unbiased optimization of Spectrometry via Closed-Loop Experimentation (MUSCLE), aims to fully automate the optimization of targeted LC-MS/MS analyses by shortening the analytical time and increasing the analytical sensitivity of targeted metabolites by iteratively suggesting new LC-MS parameters and processing the resulting peak data with a genetic algorithm [24].

2.2 Tools for isotope-assisted analysis

Stable isotopes-assisted analyses have tremendous potential in addressing the quantification issues in metabolomics studies, and they need dedicated tools to process and handle the spectral data. To this end, IsoMS processes raw data generated from LC-MS runs by performing peak picking, peak pairing, peak-pair filtering, and peak-pair intensity ratio calculations for applications in chemical isotope labeling (CIL) or isotope coded derivatization (ICD) [25]. Further, IsoMS uses a chemical derivatization method to introduce a mass tag to all metabolites that contain a common functional group (e.g., amine). Derivatization is followed by

LC-MS analysis of the labeled metabolites, enabling better detection of metabolites as well as relative quantification between labels. This data processing method is based on the use of a mass spectral feature unique to the chemical labeling approach, whereas any differential isotope-labeled metabolites are detected as co-eluting peak pairs with a fixed mass difference in the mass spectrum. A common challenge for most metabolomics software is missing data/peaks between replicates or experimental groups. IsoMS addresses the issue of missing values in a CIL LC-MS metabolomics experiment by first processing the LC-MS data set to generate metabolite ID and peak ratio information. This step is followed by a zero-fill program from MyCompoundID.org that was developed to automatically find and impute missing values by referring back to the raw LC-MS data to find the peak pair and calculate the desired intensity ratios [26]. IsoMS-Quant can be used for extracting quantitative information from a metabolomic dataset generated by CIL- LC-MS. The program reconstructs the chromatographic peaks of the light- and heavy-labeled metabolite pair to calculate the ratio of their peak areas to represent the relative concentration difference among samples [27]. Relying on chromatographic peaks to perform relative quantification, IsoMS-Quant is integrated with IsoMS for picking peak pairs and Zero-fill for retrieving missing peak pairs (see also section 2.3) in the initial peak pairs table generated by IsoMS. 'isoMETLIN' is a database developed for identification of metabolites incorporating isotopic labels, which enables users to search all computed isotopologues derived from the METLIN database on the basis of m/z values and specified isotopes of interest such as ^{13}C or ^{15}N [28]. For more confident compound identification, MS/MS spectra need to be recorded and compared against MS databases or spectra of authentic reference standards obtained under the same experimental conditions. One elegant approach to this is full, *in vivo,* stable isotopic labelling (SIL) of whole organisms. Similarly, another software tool FragExtract was developed and evaluated with LC-HRMS/MS spectra of both native ^{12}C- and uniformly ^{13}C (U-^{13}C)-labeled analytical standards to combine SIL and MS/MS for the automated interpretation of the elemental composition of fragment ions for structural elucidation of compounds [29]. Isotopologue Parameter Optimization (IPO) is a software package that aims to optimize XCMS processing parameters and can handle data from different samples and data from different methods of liquid chromatography - high resolution mass spectrometry (LC-HRMS) platforms [30]. IPO addresses this challenge by automatically evaluating the effect of a range search parameters on peak detection, retention time alignment, and peak grouping. IPO reports the

optimal configuration after targeting the maximization or minimization of important analytical results, for example, by reducing retention time differences within peak groups. IPO optimizes XCMS peak picking parameters by virtue of stable ^{13}C isotopic peaks to calculate the peak picking score and retention time correction leading to optimized grouping parameters in a controlled and reportable manner. 'Massifquant/Optimize-it' provides an optimization code and documentation for isotope trace (IT) detection in real and complex LC-MS datasets and is also integrated into XCMS [31].

2.3 Tools for handling of missing values

Missing values in metabolomics datasets are a common occurrence and originate from a number of sources, such as technical and biological reasons, and are contrivances to downstream data processing, especially statistical analysis, since many statistical calculations cannot cope with a 'zero'. To reconstruct or circumvent missing data, different forms of data imputation are a possible solution. A software application, x-VAST, deals with missing values and masking effects in data sets due to the high variation of abundant metabolites. It amends the measurement deviation enlargement, enabling rescue of low abundant masked differential metabolites [32]. An R package called metabomxtr (available through Bioconductor) was also developed to facilitate mixture-model analysis (which accounts for missing metabolites due to their absence or them being below the detection limit) in data sets where only a portion of samples have quantifiable abundance for certain metabolites [33]. Any software pipeline which addresses missing values can generate peak data based on analytical noise; therefore care must be taken to establish noise limits for each detected peak.

In MS-based metabolomics, data sets may contain missing values, which complicates some quantitative analyses, yet the presence/absence of a metabolite and a quantitative value of the abundance level of a metabolite are still possible. To tackle such challenges, a novel, kernel-based [34] score test for the metabolomics' differential expression analysis was implemented in R to capture both continuous and discrete patterns, as shown for liver cancer metabolomics data sets [35]. Furthermore, such multiple kernel–based, machine-learning approaches to predict molecular fingerprints and identify molecular structures have enabled mapping of mass spectra to molecular fingerprints, where predicted fingerprints are used to score candidate molecular structures [36].

2.4 Tools for retention time alignment and intensity drift correction

To effectively compare different metabolomics experiments, one should be confident that the same feature is compared across different data sets. For example, LC-platforms often suffer from retention time drifts within experiments, depending on chromatographic parameters such as column chemistry, mobile phase, and elution gradient. Therefore, different alignment tools were developed to group them feature across data sets together. Once a high-quality metabolomics method is established, it can still be difficult to determine the optimal parameters for peak alignment. Many peak picking tools like the already introduced XCMS have therefore options to perform retention time correction and grouping of the features as described above.

Drift effects of the peak intensity is another frequent and major source of variance in LC-MS datasets that leads to erroneous statistical results during metabolomic analyses. To this end, a methodology based on a common variance analysis before data normalization was implemented in an R package called intCor [37]. For direct matching, a peak alignment method for LC-MS data sets is available as a standalone package [38], implemented in Python, where the tool incorporates related peaks information (within LC-MS runs) and investigates their effect on alignment performance across runs [39]. A genetic programming-based approach for multiple alignment of LC-MS data (GPMS) was also available where peaks from different LC-MS maps (peak lists) are matched to allow the calculation of retention time deviation, followed by use of GP for multiple alignment of the peak lists with respect to a reference [40]. Although used in LC-DAD (diode array detector) metabolite profiling, an R package ‘alsace’ relies on a strategy based on multivariate curve resolution [41]. The tabulated results contain peak heights and areas for all features of the individual injections, and the tool allows data splitting and merging based on time-windows and parametric time warping (PTW) to align features which were shown for cassava-derived data sets as a proof-of-concept study. PTW is a highly restricted form of the warping functions that avoids over fitting, leading to alignment of peaks across samples in a robust manner by working on peak-picked features (and not on profiles). In addition, a PTW of peak lists is implemented in R [42].

PeakANOVA is another tool implemented in R that uses a hierarchical Bayesian model for finding differences between sample groups [43]. This package reduces the error of weak and non-existent covariate effects, and false-positives by using mass spectral peak data by clustering

similar peaks into latent compounds and groups which respond in a coherent way to experimental covariates. Like mentioned earlier, Table 1 provides a list of the data preprocessing and handling tools discussed above including their URLs and references.

3. Statistical tools for metabolomics data

Once the data is preprocessed and aligned, relevant mass or NMR features are normally extracted from the data using statistical approaches focusing on finding discriminative features for a certain condition. Univariate and multivariate statistical analysis techniques are used to extract relevant information from data with the aim of providing biological insights on the problem studied [44]. Further, statistical tools like the *t*-test, analysis of variance (ANOVA), principal component analysis (PCA), or partial least squares discriminant analysis (PLS-DA) are usually integral statistical components in metabolomics studies. The appropriateness of a statistical test for detecting differential expression in omics studies is determined by the data distribution. Thus, multivariate statistical analyses, such as PCA, PLS-DA, logistic regression, support vector machine, random forest; and model evaluation and validation methods such as leave-one-out cross-validation, Monte Carlo cross-validation, and receiver operating characteristic (ROC) analysis; are both finding growing applications in metabolomics research [45]. For example, SIMCA (Umetrics, Umeå, Sweden) is widely used to create PLS-DA/OPLS-DA models. [46].

DeviumWeb, short for Dynamic MultivariatE Data Analysis and VIsUalization PlatforM, is data analysis and visualization software for multivariate data [47]. DeviumWeb offers a variety of tools for metabolomic data analysis, including statistical and power analysis, data normalization, clustering, PCA, machine learning methods, and biochemical pathway enrichment analysis and visualization tools (Figure 1 A). Another noteworthy statistical analyses suite, BioStatFlow version 2.7.7., developed by National Institute of Agronomic Research, France (INRA, http://www.inra.fr/), is an extremely user friendly web-tool for analyses of -omics scale data.

Indeed, finding the causes of unwanted variation (from diverse sources ranging from operator, to instrument issues to data analyst) in metabolomics experiments are critical dimensions before proceeding to data interpretation, so as a consequence we observed the release of methods for handling this unwanted variation, and statistical approaches for the removal of

these variations to obtain normalized metabolomics data [48]. For example, to identify and quantify the contribution of relevant sources of variation in metabolomics data prior to investigation of etiological hypotheses, a Principal Component Partial R-square (PC-PR2) method combining the features of PC and multivariable linear regression analyses was developed, which has potential in large-scale, epidemiological metabolomics investigations [49]. Another example is EigenMS, a stand-alone set of two functions implemented in R which relies on a decomposition-based method for normalization of metabolomics data sets. EigenMS leaves the treatment group differences unaffected by first estimating treatment effects with an ANOVA model and bias of unknown complexity from the LC-MS metabolomics data, allowing increased sensitivity in differential analysis [50]. RepExplore is a web-service dedicated to providing more reliable and informative differential expression and abundance statistics among replicates, in addition to the analysis of a variety of –omics data, such as fully automated data processing and interactive ranking tables, whisker plots, heat maps, and PCA visualizations to interpret -omics data sets and derived statistics (Figure 2 A, B) [51]. Another open-source tool, Normalyzer, normalizes the data with 12 different normalization methods and generates a report with several quantitative and qualitative plots for comparative evaluation of different methods (Figure 2 C) [52]. Although it has proven applications in quantitative proteomics and transcriptomics, it still remains to be seen if this R package is applicable for metabolomics applications. Finally, MSPep is an R package for post-processing of metabolomic data. It performs summarization of replicates, filtering, imputation, and normalization, and it generates diagnostic plots and outputs final analytic datasets for downstream analysis [53].

In a very innovative approach to capture the micro-organismal, interspecies competition-induced, highly complex, ecological interactions-generated, untargeted metabolomics data of bioactives in ‘co-cultures’, Projected Orthogonal CHemical Encounter MONitoring (POCHEMON), written in MATLAB, was shown to be a novel multivariate data analysis method that reveals all competition-related, biochemical changes from the co-cultures, both up- or down-regulated, and *de novo* synthesized metabolites [54]. Similarly, Mutual Information-based multivariate Reporter Algorithm (MIRA) is a multivariate and combinatorial algorithm that calculates the aggregate transcriptional response around a metabolite using mutual information [55]. In this proof-of-concept study, MIRA was successful in the gene expression analysis of six knockout strains of *Escherichia coli* to help capture the metabolic dynamics of the

switch from aerobic to anaerobic respiration by eradicating statistics related issues associated with small sample sizes, analyses of genes one-by-one and as a group, and overcoming z-score associated inconsistencies.

4. Metabolite annotation and identification tools and resources

Metabolite annotation is the bridge between high quality MS data and useful biological information, be it through exploratory non-targeted profiling or targeted profiling approaches. However, identification of molecules, i.e., assignment of a chemical structure to a mass feature, is often a difficult and time-consuming process. Typically, if no authentic standard compounds are available, but fragmentation data (i.e., MS2 or MS^n data) is acquired, the compounds are putatively identified by comparing the measured MS/MS fragments against spectra of tandem MS databases that are publicly available. MassBank [56, 57], METLIN [58, 59], or NIST MS/MS [60] databases are commonly used for spectral comparisons. However, there are substantial issues in database based annotation process, such as completeness of the databases, conversion from raw data to molecular structure, and false discovery rate of matched hits, among others. To solve these limitations, several software tools were published in recent years, many of which attempt to use the structural information that MS fragmentation spectra generate of selected metabolites. Furthermore, spectral and compound databases have become indispensable resources for metabolite annotation and identification. Unsurprisingly, several tools were released that make use of the increasing power and coverage of metabolite and chemical compound databases. The diversity of tools is such that it is hard to make a classification within these metabolite annotation and identification tools, so the authors decided to present them alphabetically here.

BioSM is a chemiinformatics tool that uses known mammalian biochemical compounds (such as scaffolds) and graph matching methods to identify biochemical structures in chemical structure space for qualitative and quantitative biochemical annotation, leading to the identification of unknown biochemical structures in metabolomics [61]. BioSMXpress is an advanced BioSM that provides enormous resources for identification of unknown biochemical structures in metabolomics, based on matching graph methods against known biochemical structures in chemical structure space [62].

The recently introduced competitive fragmentation modeling (CFM), is a machine learning approach which finds application in both a MS/MS spectrum prediction task (i.e., predicting the mass spectrum from a chemical structure) and in a putative metabolite identification task (ranking possible structures for a target MS/MS spectrum) [63]. CFM-ID is the freely available web server based on the CFM approach aiding in the interpretation of tandem mass spectra (MS/MS) for the purpose of automated metabolite identification by facilitating annotation of the peaks in a spectrum for a known chemical structure, prediction of spectra for a given chemical structure, and putative metabolite identification [64]. Another tool, ChemProphet, can be used to annotate compounds using databases such as PubChem and ChemSpider for LC-HRMS with multistage, CID–provided spectral information on conjugated metabolites [65]. The annotation of conjugates involves recognition of the type and number of conjugates, compound search and annotation of the deconjugated form, and i*n silico* evaluation of the candidate conjugate, where spectra are assigned to each candidate by automatically exploring the substructures corresponding to the observed product ion spectrum.

GS-align was developed for glycan structure alignment and similarity measurement, in addition to allowing template-based glycan structure prediction [66]. Cosmiq is a tool (available on Bioconductor) for preprocessing LC-MS and GC-MS data, with a focus on metabolomics or lipidomics applications, which aids in the detection of low abundant signals, as it generates master maps of the m/z and retention data from all acquired runs before generating a peak detection algorithm, leading to robust identification and quantification of low-intensity MS signals compared to conventional approaches [67].

Feature Credentialing is a tool available as an R package, a simple software algorithm that is freely available and has been validated with reversed-phase and hydrophilic interaction LC as well as Agilent, ThermoScientific, AB SCIEX, and LECO mass spectrometers [68]. Shown with the help of *E. coli* extracts in a proof-of-concept study that accurately compared the number of non-artifactual features yielded by different experimental approaches (i.e., cells cultured in regular and ^{13}C-enriched media and on the basis of isotope spacing and intensity), the credentialing platform can readily be applied by any laboratory to optimize their untargeted metabolomic pipeline for metabolite extraction, chromatographic separation, mass spectrometric detection, and bioinformatics processing.

HAMMER [69] is a software package for automation of Mass Frontier™ (ThermoScientific) analysis, a commercial software, allowing high throughput automation and generation of *in silico* mass spectral libraries and performing spectral matching [70].

HR3 is a software program for annotation of elemental composition of molecular ions detected in MS, based on improved filtering rules followed by ultrafast querying in publicly available compound databases (i.e., PubChem, a general source of 1.3 million unique chemical formulas and a plant metabolomics database containing ca. 100, 000 formulas), which can potentially be used as a source of naturally occurring compounds [71].

In the MS (E) mode, a mass-spectrometer fragments all molecular ions generated from a sample allowing to record time-resolved molecular ion and fragment ion data for both molecular ions and fragments. For lipids (lipidomics experiments), the freely available software tool LipidPro matches the retention time-aligned m/z data of molecular- and fragment ions with a lipid database and generates a report on all identified lipids for large datasets generated by LC-MS/MS that uses the advanced, data-independent acquisition (DIA) mode MS (E) [72].

MS Annotation, based on *in silico* Generated Metabolites (MAGMa), is an automatic annotation software of MS/MS-based fragmentation data [73, 74] that can be used through a web server. The MS^n or MS/MS data are uploadable as a hierarchical tree of fragment peaks based on m/z values, while after submitting the data candidate molecules are automatically retrieved from PubChem, KEGG, or HMDB. Recently, the MAGMa approach was combined with *in silico* predictions of human metabolized green tea compounds, which resulted in the successful annotation of 100 urinary metabolites of green tea drinkers [75].

MAIT (Metabolite Automatic Identification Toolkit) is implemented in R for LC-MS based metabolomics data set analysis to generate statistical results, peak annotation, and metabolite identification as outputs, with a focus on the peak annotation stage [76]. Metabolome Searcher is a web-based tool for putative compound identification of MS output based on genome-restricted metabolic capability, by directly searching genome-constructed metabolic databases, especially for identifications of large data sets for metabolites that are not in compound databases [77].

METabolite COmpound Feature Extraction and Annotation (MET-COFEA) detects and clusters chromatographic peak features for each metabolite compound by comprehensive evaluation of retention time and peak shape criteria, followed by annotations of the associations between each peak's observed m/z value with the corresponding metabolite compound's molecular mass [78]. MetAssign, implemented as part of the mzMatch metabolomics analysis pipeline, is based on Bayesian modeling that combines information from the mass-to-charge ratio, retention time, and intensity of each peak, with increased accuracy of peak annotation [79].

MS2Analyzer enables user-defined searches of thousands of spectra for mass spectral features such as neutral losses, m/z differences, and product and precursor ions from MS/MS spectra [80]. A proof-of-concept validation study showed that 92.1 ± 6.4% of the 13 typical neutral losses (i.e., acetylations, cysteine conjugates, or glycosylations) correctly annotated the associated substructures of complex lipids in microalgae. Metabolite Identification via Database Searching (MIDAS) is an open source algorithm developed to evaluate a metabolite-spectrum match (MSM), where it first enlists fragments from a metabolite-by-bond dissociation, calculates the plausibility of fragments based on their fragmentation patterns, and then scores MSM to assess the match quality of experimental MS/MS spectrum from collision-induced dissociation (CID) against a predicted CID MS/MS spectrum [81].

The mzCloud database [82] has become a powerful computational and database framework for compound annotation and identification for LC-MS^n and LC-MS/MS data sets, having both collision induced dissociation (CID) and higher-energy collision dissociation (HCD) fragmentation spectra available for more than 3000 unique molecules (Figure 3). Moreover, it assists analysts in identifying compounds, even when they are not present in the library, using substructure search.

Extending the Biochemical Network Integrated Computational Explorer (BNICE) framework, Metabolic *In silico* Network Expansions (MINEs) expand curated biochemistry databases by considering the products of known enzymatic activities to address metabolite identifications from spectral features [83]. Using expert-curated reaction rules based on the Enzyme Commission classification system to propose novel chemical structures and reactions, 198 of these generalized chemical reaction rules were applied once to all compounds in KEGG, EcoCyc, and the YMDB, resulting in MINE databases of predicted products and chemical

reactions. The KEGG Compound database, consisting of over 571,000 compounds, and a proof-of-concept study MINE were used together to annotate untargeted metabolomics data from an *E. coli* knockout study.

Another noteworthy tool is the 'Molecular Networking' approach that allows MS/MS data visualization of complex metabolomics data sets to aid in their visualization and interpretation. Thus, molecular networking combines metabolomes into a single network in which the MS/MS data is organized based on chemical similarity; this was applied to MS/MS data sets from diverse samples [84]. In doing so, it was shown that molecular networking is a powerful complement to traditional dereplication strategies, whereby including structurally known synthetic structures in the network analysis leads to the discovery of novel natural products, as shown in microbial samples [85, 86]. In an integrated strategy, by combining i) high scan-speed QTOF, coupled with ultra-high-pressure liquid chromatography (UHPLC) for data collection, with ii) molecular networking as an organizational tool, iii) Cytoscape, spatial mapping as visualization tool, iv) and an automated database search for rapid identification of several classes of metabolites including lipids, several discriminative metabolites were identified from direct tissue extractions of human lung associated with cystic fibrosis [87].

Mass Spectrometry-Data Independent AnaLysis software (MS-DIAL) is an open source software pipeline that was developed for data-independent, acquisition-based identification and quantification of small molecules by mass spectral deconvolution. MS-DIAL was applied to the analysis and identification of chemotaxonomic grouping in nine algal strains based on identification of 1,023 lipids by reversed-phase LC-MS/MS analysis [88]. MSFINDER is the first software program to predict molecular structures from experimental MS and MS/MS spectra that extensively utilize online resources [89]. The source code is written in the C and predicts the molecular skeleton, i.e., the first block of InChIKey.

The software tool 'mzGroupAnalyzer' is an algorithm to automatically explore the metabolome for the detection of metabolite transformations caused by biochemical or chemical modifications, whereby pathways are extracted directly from the data and putative novel structures can be identified [90]. The tool mzGroupAnalyzer is integrated into the graphical user interface (GUI) of the COVAIN toolbox. ProbMetab is an R package that follows a Bayesian

model for automated probabilistic LC-MS metabolome annotation [91]. RAMClustR is an R package that groups signals from MS data into spectra without relying on the predictability of the in-source phenomenon and the annotation of MS signals by incorporating indiscriminant MS/MS (idMS/MS) data, leading to the identification of metabolites from a single experiment [92].

Finally, among other promising approaches used for mining LC-MS-based metabolite profiling data for specific metabolite classes is achieved by calculating relative mass defects (RMDs) from molecular and fragment ions [93]. As shown in this proof-of-concept study for sesquiterpenoids of wild tomato *Solanum habrochaites* LA1777 trichomes, this strategy enabled identification of compounds in complex plant mixtures independent of retention time, abundance, and elemental formula and led to the prediction of 24 novel elemental formulas corresponding to glycosylated sesquiterpenoid metabolites and 200 distinct sesquiterpenoid metabolites. Here, it is interesting to mention that chemoselective (CS) probes that tag to metabolite functional groups when combined with high mass accuracy provide aids in metabolite identification and quantification [94], as well. Further, using a novel algorithm that efficiently detects functional groups within existing metabolite databases such as KEGG and HMDB allows for combined molecular formula and functional group queries to aid in metabolite identification without *a priori* knowledge. *In silico* analysis of CS-tagging strategies demonstrated that combined FT-MS derived molecular formulae and CS-tagging can uniquely identify up to 71% of KEGG and 37% of the combined KEGG/HMDB database. Table 1 lists the available annotation resources.

5. Pathway analysis, network construction, visualization, and biological interpretation tools

Availability, construction, re-construction, and visualization of metabolic pathways as well as biochemical networks constitute the most important steps toward biological interpretation of metabolomics datasets. For example, computational tools can use available biochemical pathway information to rank putative annotations based on the number of metabolites measured within a given biochemical pathway. In bioinformatics, visualization provides researchers with an overview of large and complex datasets, assists in discovering or inferring patterns and relationships within data sets, for facilitating data mining, drawing conclusions, hypothesis generation, and rational interpretation of results. Further, the need for tools for storage, query,

browsing, analysis, and visualization has been critical for software development in this area. Unsurprisingly, we have witnessed the release of numerous 'pathway visualization tools' in recent years, of which we list several here.

5.1 Tools for pathway-scale mapping and visualizations

PathWhiz is a web server that generates colorful, aesthetic, and biologically meaningful pathways that are both machine-readable and are interactive [95]. Further, it readily generates complex pathways by using a specially designed drawing tool to quickly render metabolites (including automated structure generation), proteins, nucleic acids, membranes, subcellular structures, cells, tissues, and organs, thus allowing both small-molecule and protein/gene pathways constructions. It also combines multiple pathway processes such as reactions, interactions, binding events, and transport activities as well. TrackSM, developed in MATLAB, is a chemoinformatics tool used to associate chemical compounds to metabolic pathways based on molecular structure-matching methods [96]. TrackSM successfully associated 93% of tested structures to their correct KEGG pathway class and 88% to correct individual KEGG pathway. MarVis-Pathway, integrated to MarVis-Suite, is a tool that allows identification of metabolites by annotation of pathways from cross-omics data, pathway enrichment, and meta-analysis, thus aiding in mapping of data set features by ID, name, and accurate mass (including adduct and isotope correction information) from MS Data [97]. InCroMAP is a data integration, analysis, and visualization tool for metabolomics, and other multi-omics data sets for integrated enrichment analysis and pathway-based visualization [98]. Similarly, integrative Pathway Enrichment Analysis Platform (iPEAP) allows analyses of multi-omics data sets including metabolomics, and GWAS data and aggregates the pathway enrichment results generated in different high-throughput experiments quantitatively for the integration, comparison, and evaluation of diverse data types [99]. Kpath is another database for integration of information related to metabolic pathways, provides a navigational interface for browsing and use of the integrated data to build metabolic networks based on existing knowledge [100]. Pathomx is an open source cross-platform analysis tool developed in Python for import, processing and visualization of (all types of metabolomics) data alongside back end integration with MATLAB and R [101]. In a proof-of-concept study, the application of the software to 1D and 2D ^{1}H NMR

metabolomics data analysis for mammalian cell growth under hypoxic conditions was also provided. Similarly, R-based tools such as pwOmics are now available which perform pathway-based, level-specific, time-series data comparison of multiple -omics using public database knowledge [102]. In a cross-platform consensus analysis, the combined interpretation of regulatory effects over time was observed via network reconstruction and inference methods, leading to consensus graphical networks (see also section 5.2). However, as for all pathway visualization tools, the application of pwOmics to metabolomics datasets needs to be carefully conducted and checked, with the obvious challenges in ID conversion and mapping issues associated with metabolites.

5.2 Tools for network-scale mapping and visualizations

Genome-scale metabolic networks have caught the attention of biological researchers, as they assist in adding insights into biological questions, as compared to the much more simplistic view of one-dimensional biochemical pathways such as KEGG. PathCase Metabolomics Analysis Workbench (PathCaseMAW) runs on a manually-created, generic, mammalian metabolic network, is a database-enabled framework (with its user-friendly interface), and can be used to generate new metabolic networks and/or update an existing metabolic network [55]. These genome-scale metabolic networks are accessible through a web interface or an iPad application that implement an integrated steady-state metabolic network dynamics analysis (SMDA) algorithm. MetaMapR leverages the KEGG [103] and PubChem [104] databases to provide methods for integration and visualization of complex metabolomics experiments, even in cases where biochemical domain knowledge or molecular annotations are unknown [105]. For example, MetaMapR has been used to integrate both biochemical reaction information with molecular structural and mass spectral similarity to identify pathway-independent relationships among metabolites, even in cases where the metabolite annotation or structural identify was unknown [106-108]. Network calculation is further bolstered through an interface to the Chemical Translation System [109], allowing metabolite identifier translation between > 200 common biochemical databases. On a personal note, the authors find MetaMapR extremely useful and easy to use when combining a wide range of data sets, grouping them according to

both chemical structure and functionality, and providing biological insights (Figure 1 B). An example is provided in Figure 1 from another recent review manuscript. [110].

Metabnet is another R package for metabolic network analysis based on a high-resolution targeted, metabolome-wide association study (MWAS) of specific metabolites for metabolic pathways and network structures prediction [111]. Several other tools were developed such as additional functionality, applications (apps), plugins, or add-ons to Cytoscape [112] for network analysis such as CentiScaPe [113], KEGGscape [114], GeneMANIA [115], MetDisease [116, 117] and, recently discussed elsewhere [118], KeyPathwayMiner 4.0 [119] for integration of multiomics data sets. A huge number of other similar tools, which are not introduced here, continue to pile up. Another noteworthy approach for prediction of enzymatic reaction networks from a metabolome-scale compound set and finding chemical substructure transformation patterns in multistep reaction sequences is a recursive, supervised approach [120]. MetaNET is an open source, platform-independent, and web-accessible resource for metabolic network analysis through flux balance and variability, chemical species participation, cycles, and extreme paths identification [121]. This web tool is implemented in Galaxy workflow and has Systems Biology Research Tool (SBRT) among its components. FunRich, is an open access, standalone functional enrichment and network analysis tool that performs functional enrichment analysis on background taxonomic and custom-made databases that are integrated from heterogeneous omics and metabolomics datasets [122]. Similarly, Network Portal serves as a modular database for the integration of custom and public data sets, with inference algorithms and tools for the storage, visualization, and analysis of biological networks [123]. This portal is integrated into the Gaggle framework alongside social networking capabilities for collaborative projects. Presently, the database contains networks for 13 prokaryotes from diverse phylogenetic clades (4678 co-regulated gene modules, 3466 regulators, and 9291 cis-regulatory motifs). A chemical graph alignment algorithm, PACHA (Pairwise Chemical Aligner) detects the regioisomer-sensitive connectivity between the aligned substructures of two compounds, thus rendering it useful for reaction annotation that assigns potential reaction characteristics such as EC (Enzyme Commission) numbers and PIERO (Enzymatic Reaction Ontology for Partial Information) terms to substrate-product pairs [124]. SimIndex (SI) and SimZyme, implemented in Python, use the chemical similarity of 2D chemical fingerprints to efficiently navigate large metabolic networks

and propose enzymatic connections [125]. This study also reported a Byers–Waterman type pathway search algorithm for further paring down pertinent networks where general graph search algorithms are extremely slow, for instance, for very extensive metabolic networks.

5.3 Other useful small molecule tools and approaches that involve pathways or enzymes

BiNChE is an open-source enrichment analysis tool for small molecules based on their Chemical Entities of Biological Interest (ChEBI) role or similar structural ontologies that displays interactive graphs that are exportable in high resolution image and network formats [126]. This may aid in the exploration of large sets of metabolites identified in metabolomics or other systems biology research contexts. In the PlantSEED server, a significant number of genome-scale metabolisms for different organisms were reconstructed and made available [127]. The Selective Paired Ion Contrast (SPICA) algorithm extracts biologically relevant information from the noisiest of metabolomic data sets, as it relies on analyzing ion-pairs by exhaustively considering all possible ion-pair combinations and statistical comparisons between sample groups, leading to biomarker discovery using support vector machines (SVMs) [128]. KiMoSys is web-based and integrates public experimental and published data and associated model repository with computational tools, providing a novel application for facilitating data storage and sharing for the systems biology community [129]. This web application, implemented using the Ruby on Rails framework, is freely available and contains concentration data of metabolites and enzymes and flux data. Integrated Interactome System (IIS) is a web-based, free, and integrative platform for annotation, analysis, and visualization of the interaction profiles of proteins/genes, metabolites, and drugs of interest, as well as metabolomics data sets [130]. It consists of submission, search, annotation, and interactome modules for building networks that gather novel identified interactions, protein, and metabolite expression/concentration levels, subcellular localization and computed topological metrics, GO biological processes, and KEGG pathways enrichment. TAPIR is a Python visualization software for chromatograms and peaks identified in targeted proteomics and metabolomics experiments [131]. The input formats are mzML for raw data storage and TraML for encoding the hierarchical relationships between transitions, thus aiding in visualization of high-throughput –omics datasets. For ‘interactomics’, other efforts include multivariate gene-metabolome association studies using a Bayesian-

reduced, rank regression approach [132]. Table 1 lists these pathway, network, and interpretation tools.

6. Metabolomics libraries, databases, experiment repositories, and meta-data storage

In addition to data handling, processing, and annotation tools, several chemoinformatics resources in the form of libraries, databases, and applications/ add-ons were released or extensively updated during this two year period. Publically-accessible, large databases aiding metabolomics discovery include PubChem [133] and ChemSpider [134] which contain organic and biological molecules that are searchable by exact mass or compound name, whereas METLIN [135], MassBank, and MzCloud contain curated compounds of biological origin that are searchable by exact mass and fragmentation spectrum. It is noteworthy that the commercial NIST database now not only contains a large body of spectra relevant for GC-MS (EI fragmentation), but also MS/MS spectra relevant for LC-MS (ESI fragmentation) to aid in metabolite annotations. KNApSAcK Family Databases holds potential to find applications in numerous metabolomics approaches ranging from natural products–based drug discovery to animal, plant, and microbial metabolomics [136, 137]. These databases also facilitated in the development of network-based approaches to analyze relationships between 3D structure and biological activity of metabolites. In a proof-of-concept study using a data set consisting of 2072 secondary metabolites and 140 biological activities reported in KNApSAcK Metabolite Activity DB, about 983 statistically significant structure group-biological activity pairs were obtained.

6.1 Databases useful for specialized applications

Databases help in detection, identification, classification and interrogation of molecular structures to help us interpret their relevance to biological phenomena. Easier navigation, integration and extraction of information out of the chemical and biological databases from diverse fields of study would enhance the usefulness to metabolomics research. Within the past two years, numerous new databases have surfaced, with applications ranging from plant and microbial to pharmaco-metabolomics. Tea Metabolome database (TMDB) is a manually curated and web-accessible database of metabolites found in tea (*Camellia sinesis*) and related species which is a repository of 1393 metabolites [138]. Each metabolite entry in TMDB contains an average of 24 separate data fields, such as compound structure, formula, molecular weight,

name, CAS registry number, compound type, health benefits, reference literature, NMR, MS data, and the corresponding IDs from databases such as HMDB and PubMed. BioPhytMol database is a drug-discovery community resource and database with anti-mycobacterial phytomolecules and plant extracts that holds 2582 entries including collective information for188 plant families (for 692 genera and 808 species) from global flora, that were manually curated from literature including 633 phytomolecules (with structures) curated against 25 target mycobacteria [139]. The ESSential OIL DataBase (EssOilDB) is an electronic database designed to provide knowledge resource for plant essential oils that allows user queries on volatile profiles of native, invasive, normal or stressed plants across taxonomic clades, geographical locations, and several other biotic and abiotic influences [140]. EssOilDB boasts 123 041 essential oil records, spanning a century of published reports on volatile profiles with data from 92 plant taxonomic families. Microbial VOCs (mVOCs) is an up-to-date collection on microbial volatile organic compounds (VOCs) [141]. These latter two would prove to be important resources for essential oil and volatiles-based metabolomics. Environmental metabolomics is defined as the use of metabolomics techniques to characterize the metabolic response of organisms to natural and anthropogenic stressors in the environment [142]. To this end, for applicability in marine environment research, a boutique database was developed for efficient data analysis and selection of mass spectral targets for metabolite identification and has led to the development of 'domdb' [143]. PhenoMeter, is a metabolomics database search server that accepts metabolite response patterns as queries and searches the MetaPhen database of reference patterns for statistically significantly patterns towards deciphering functional links, where the assigned PhenoMeter Score (PM Score), as a function of both Pearson correlation and Fisher's exact test of directional overlap, reveals insights on environmental and genetic perturbations in the metabolome [144]. Exposome is defined as the totality of all human environmental exposures from conception to death. The Toxin-Toxin-Target Database (T3DB), designed to capture information about the toxic exposome [145], includes compounds (>3600), targets (>2000) and gene expression datasets (>15 000 genes), chemical ontologies, and a large number of referential NMR, MS/MS and GC-MS spectra of human exposome. SwissLipids provides curated knowledge of lipid structures and metabolism used to generate an *in silico* library of feasible lipid structures [146]. A hierarchical classification links MS analytical outputs to all possible lipid structures, metabolic reactions, and enzymes and provides a reference namespace for

lipidomic data publication, data exploration, and hypothesis generation. The present version contains ~244 000 known and theoretically possible lipid structures, ~800 proteins, and curated links ~620 published peer-reviewed publications.

6.2 Databases and tools for metabolomics data management, archiving, and sharing

For metabolomics-scale, 'big data' handling, dissemination, and archiving, although the role of GigaDB [147], the European COordination of Standards in MetabOlomicS (COSMOS) consortium [148], and MetaboLights [149] as front runners is well known. MetaboLights is the first comprehensive, cross-species, cross-platform metabolomics database that can hold complete metabolomics experiments. Since its launch, more of such repositories were launched, i.e., the Metabolomics Workbench [150], and the MetabolomeXchange [151] website was created to provide an overview. The available datasets [152] can now serve as a benchmark for direct-infusion MS (DIMS) metabolomics, which were derived using best-practice workflows and rigorous quality assessments. In addition, the Investigation/ Study/ Design (ISA) ISA-Tab format is a metadata standard that has gained a lot of momentum since first being released in 2008. It is a source metadata tracking framework that facilitates standards-compliant collection, curation, visualization, storage, and sharing of datasets for analysis and publication [153]. Recently, linkedISA [154] for the semantic interoperability and the Risa package [155] that supports the ISA format by integration with R have extended the way MS-based metabolomics data sets are handled for meeting the metabolomics standards initiative (MSI) [156-160]. Metabolonote is a wiki-based database for managing hierarchical metadata of metabolome analyses [161], wherein the Togo Metabolome Data (TogoMD) data format was introduced with an ID system that is required for unique access to each level of the tree-structured metadata, such as study purpose, sample, analytical method, and data analysis. Presently, Metabolonote houses 808 metadata obtained from 35 biological species. KOMICS (The Kazusa Metabolomics Portal) consists of free tools and databases for preprocessing, mining, visualization, and publication of metabolomics data [162], and is strengthened with additional tools, i.e., PowerGet and FragmentAlign for manual curation function for the results of metabolite feature alignments. MetaDB is an open-source web application for metabolomics metadata management and data processing where the analysis of untargeted data is done using the R package MetaMS [163,

164]. Laboratory Information Management System (LIMS) has changed the way research and development, data generation and archiving, and project management are perceived presently. Quality and TRacEability Data System (QTREDS) is a software platform written in the Ruby programming language and developed using the Rails framework with visualizations that are JavaScript based and is freely available to academic users for management of 'omics' laboratories [165]. In addition, MASTR-MS, developed by Metabolomics Australia and Australian Bioinformatics Facility as part of the Bioplatforms Australia national initiative, is an online LIMS solution for metabolomics that manages researchers' needs, from experimental design to the capture and management of raw and processed data of metabolomics experiments [166, 167]. Similarly, Yabi is a software system that is adaptable to a range of execution and data environments in an open source implementation, and it provides an analysis workflow environment that can create and reuse workflows for a range of applications in metabolomics including quality assessment and statistical analyses [168, 169].

6.3 Others tools relevant to metabolomics databases and libraries

Currently, ProteoWizard library [170] includes the msConvert tool which is able to read data files from different mass spectrometers and convert them into open data formats that can be used to store and exchange MS data. The file format mzTAB is a standard flexible, tab-delimited format proposed by Proteomics Standards Initiative (PSI) which can capture identification and quantification results coming from MS-based proteomics [171], and may have scopes for application in metabolomics approaches. An open-source Java application programming interface for mzTab, i.e., jmzTab [172] is integrated and used in LipidDataAnalyzer [173]. pyOpenMS is an open-source, Python-based interface to the C^{++} OpenMS library, providing access to a feature-rich, open-source algorithm library for MS-based analysis by allowing access to the data structures and algorithms implemented in OpenMS, specifically those for file access (mzXML, mzML, TraML, mzIdentML etc.), basic signal processing (smoothing, filtering, de-isotoping, and peak-picking) and complex data analysis for proteomics applications [174].

PolySearch2 is an online text-mining system for exploring relationships between biomedical entities such as human diseases, genes, SNPs, proteins, drugs, metabolites, toxins, metabolic pathways, organs, tissues, subcellular organelles, positive health effects, negative health effects, drug actions, Gene Ontology terms, MeSH terms, ICD-10 medical codes,

biological taxonomies, and chemical taxonomies [175]. To aid in spectral visualization from MS, IR and NMR, SpeckTackle, a custom-tailored cross-browser compatible JavaScript charting library for spectroscopy was developed [176]. BioMet Toolbox 2.0 enables the user to work with biological data in a web user interface using tools for constructing metabolic pathways and other omics datasets [177].

In addition, MS-imaging technology is upcoming and the number of tools that can handle imaging data is rising, for example for plant metabolomics [178]. Metabolite Imager is a free, Java-based metabolomics application that enables customized analysis and visualization of the metabolite distributions in tissues acquired through MS-based imaging approaches [179]. Metabolite Imager algorithms perform customized targeted searching of metabolites through user-defined and publicly-available databases, enabling the analysis of spatial distributions of large metabolite numbers in tissue sections. Metabolite Imager's automated, 2D image generator functionality has several customizable features for producing HR images. To assist in Matrix Assisted Laser Desorption Ionization-Imaging Mass Spectrometry (MALDI-IMS) based metabolomics data acquisition for obtaining spatial distribution of molecules in samples, an algorithm, EXIMS, was used [180]. An unsupervised algorithm uses Sliding Window Normalization (SWN) and a new spatial distribution based peak picking method developed based on Gray level Co-Occurrence (GCO) matrices, followed by clustering of metabolites to extract features from molecular images as shown for a proof-of-concept metabolomics dataset of a *Eucalyptus sps*. leaf extracts. In addition, the algorithm was shown to be performing better than the existent ones, as well as applicable to proteomics-based imaging.

7. GC-MS-based tools

During this period, several important tools have surfaced to aid in the analysis and interpretation of GC-MS-based metabolomics data. Nonetheless, compared to LC-MS-based tools, their numbers seem to be fewer. For instance, Maui-VIA allows users to process, inspect, and correct large numbers of GC–MS samples, and provides functionalities for retention index calculation, targeted library search, visual annotation, alignment, correction interface, and metabolite quantification [181]. Another recent study showed the capability of orthogonal signal deconvolution (OSD), a novel algorithm based on blind source separation, to extract the spectra

of compounds appearing in comprehensive gas chromatography - MS (GC x GC-MS) analyses for metabolomics applications [182].

MetaMS (available in Bioconductor) is an open-source software written in R based on XCMS [163] and CAMERA [183] pipelines for untargeted metabolomics for GC–MS and relies on the estimation of relative concentrations of compounds and metabolite identification using in-house databases [163], thus promising immense benefit for untargeted metabolomics by allowing screening of in-house databases of chemical standards. PScore is a platform independent package that scores peaks based on their likelihood of representing metabolites defined in a MS library, and is implemented in an R package called MetaBox [184]. This package allows reliable identification and quantification of metabolites analyzed by GC-MS. BiPACE 2D, available with the Maltcms framework, is an automated algorithm for retention time alignment of peaks from 2D GC-MS experiments and evaluates them on published datasets using the mSPA, SWPA, and Guineu algorithms [185]. When compared with AMDIS (Automated Mass Spectral Deconvolution and Identification System), MetaBox reported lower percentages of false positives and false negatives in a volatile organic compound analysis. The tools discussed above are provided in Table 1.

8. NMR-based tools

NMR spectroscopy is routinely applied for metabolomics applications without purification, where overlapping NMR peaks offer tremendous challenges for the comprehensive and accurate identification of metabolites. In addition, the application of NMR in metabolomics is delimited by its inherent low sensitivity. Given the role of NMR in unknown discovery and structural elucidation, it is imperative to address the growth of NMR-based metabolomics tools in a dedicated section. As observed for MS-based metabolomics, spectral annotation and assignment of NMR data remains a major challenge as well. To this end, pattern recognition based approaches such as the automated assignment method were developed, based on matching the pattern of peaks in the spectrum rather than absolute tolerance thresholds using a combination of geometric hashing and similarity scoring techniques, rely on spectral calibration using internal or external standards [186]. Tests using 719 metabolites in the HMDB correctly annotated 100% of the metabolites in a spectrum, even under poor data conditions such as 50% missing peaks, large deviations in chemical shifts, and overwhelming data redundancy. ‘Focus’

is an integrated, open-source software program that provides a complete data analysis workflow for 1D NMR-based metabolomics [187]. Focus allows users to obtain a NMR peak feature matrix and metabolite identification scores, whereas a new spectral alignment algorithm, RUNAS, allows peak alignment without a reference spectrum. Normalization of the Bayesian automated metabolite analyzer for NMR (BATMAN) is an R package, which performs automated metabolite deconvolution and quantification from complex NMR spectra, aiding in analyses of a large numbers of spectra [188].

Nonetheless, identification becomes cumbersome in case of samples carrying mixtures of unknown origin. Metabnorm is another tool utilizing a normalization approach based on a mixed model, with simultaneous estimation of a correlation matrix for NMR-based metabolomics data [189]. Bayesil is an online tool that can determine a metabolic profile, for example from 1D ^{1}H NMR spectrum of a complex biofluid, without any human guidance [190]. Bayesil performs all of the required spectral processing steps (i.e., Fourier transformation, phasing, solvent-removal, chemical shift referencing, baseline correction, line shape convolution), followed by matching the resulting spectrum against a reference compound library that contains the signatures of individual metabolites (Figure 4). SENECA is a package for Computer Assisted Structure Elucidation (CASE) of organic molecules that uses 1D and 2D NMR spectroscopy for the CASE process, aiding in spectroscopically underdetermined structure elucidation problems and in new feature prediction, based on natural product-likeness, allowing better ranking of the correct structure prediction [191]. Complex Mixture Analysis by NMR (COLMAR) is a database and query algorithm for the analysis of ^{13}C–^{1}H HSQC spectra that unifies NMR spectroscopic information on ~555 metabolites from both the Biological Magnetic Resonance Data Bank (BMRB) and HMDB [192]. Similarly, ^{1}H(^{13}C)-TOCCATA is a customized database containing complete ^{1}H and ^{13}C chemical shift information on individual spin systems and isomeric states of common metabolites as they directly correspond to cross sections of 2D ^{1}H–^{1}H TOCSY and 2D ^{13}C–^{1}H HSQC-TOCSY spectra, allowing the straightforward and unambiguous identification of metabolites of complex metabolic mixtures based on a collection of ~455 metabolites [193]. The mQTL.NMR package (available at Bioconductor), written in R, aids in NMR-based, untargeted metabolomic data processing, quantitative analysis, and genetic mapping [194]. The package performs preprocessing of collinear NMR data sets to reduce the multiple testing

burden, generates accurate mQTL mapping in human and rodent models, statistically improves structural assignment of genetically determined metabolites, and illustrates data with visualization tools.

MVAPACK is an open-source platform for complete NMR metabolomics data handling which aids users in loading and visualization, phasing and referencing, binning, alignment, and multivariate statistical analyses for NMR-based metabolomics data sets [195]. During the revision of this compilation, the release of another interesting tool, ChemoSpec [196] was observed, which is available to metabolomics researchers for exploratory chemometrics for spectroscopy-generated metabolomics datasets such as NMR, infrared or Raman spectroscopes for spectral inspection and alignment, HCA, and PCA analyses, as well as for finer analysis such as statistical total correlation spectroscopy (STOCSY) analysis.

Another novel approach called 'NMR/MS Translator' is a fully automated approach that synergistically combines the power of NMR and MS, with enhanced accuracy and efficiency for the identification of metabolites [197]. In this approach, metabolite candidates are identified from 1D or 2D NMR spectra by NMR database query, followed by the determination of the masses (m/z) of their possible ions, adducts, fragments, and isotope distributions. Common metabolites are identified by comparing the MS1 spectrum to NMR spectra with enhanced confidence, and a validation of the NMR-derived metabolites as was shown for the urine metabolome in this proof-of-concept study. Recently, two-dimensional ^{13}C-^{13}C correlation experiments like INADEQUATE (incredible natural abundance double quantum transfer experiment) have been proposed for structural elucidation of natural products and metabolomics analysis [198]. In addition, a semi-automated approach called INETA (INADEQUATE network analysis) for the untargeted analysis of INADEQUATE data sets using an in silico INADEQUATE database was demonstrated using isotopically labeled *Caenorhabditis elegans* mixtures as a proof-of-concept study. The tools discussed above are listed in Table 1.

9. Multifunctional tools

Workflows that are free, web-based, user friendly, and provide holistic solutions in terms of data processing, annotation, and statistical analyses, followed by biological interpretation, are growing in popularity. In this review, we term the tool multifunctional if a user can enter these

software suites with raw data (or mzML/mzXML) files and run through all the necessary processing, statistical, annotation, and pathway mapping steps using just a few clicks of the mouse and some basic knowledge about the steps for decision making out of the available options. It is noteworthy that many of these tools can handle different types of *omics* data, though LC-MS type of data is the most accepted one.

An updated MetaboAnalyst, version 3.0 [199], a popular web server available for comprehensive metabolomic data analysis, visualization, and interpretation, was released with improved visualization, performance, capacity, user interactivity, and added modules for biomarker analysis, power analysis for improved planning of metabolomics studies, and integrative pathway analysis for both genes and metabolites [200]. XCMS Online, developed by Scripps Center for Metabolomics and Mass Spectrometry (http://masspec.scripps.edu/), is a cloud-based informatics platform designed to process and visualize mass-spectrometry-based, untargeted metabolomic data for dependent (paired), two-group comparisons, meta-analysis, and multigroup comparisons, with comprehensive statistical outputs and interactive uni- and multi-variate visualization tools (cloud and PCA plots) in real time by adjusting the threshold and range of various parameters [201]. Furthermore, autonomous metabolomic workflows were established for combining MS analysis with MS/MS data acquisition [202] and were designed to allow for simultaneous data processing and metabolite characterization using combined resources XCMS and METLIN [58, 59, 135].

Mass++, available from MassBank is a plug-in style visualization and analysis tool for MS-based metabolomics with a huge number of plug-ins for visualization, data pre-processing, and annotation of data sets [203]. The plug-ins support all major vendors' raw data formats in an interchangeable manner and work as both a desktop tool and a software development platform. Original functions can be developed without editing the Mass++ source code. The author finds that particular potential and usefulness of Mass++ in metabolomics is owed to its versatility in handling file formats in a vendor-independent manner and its ability to interconvert file formats, visualizations, and press-button annotations at MassBank. Along with providing access to label-free quantitation in metabolomics, this freely available tool also holds potential in proteomics. Nevertheless, its application for handling of metabolomics data sets in the truest sense--in terms of publications, remains to be seen.

MASSyPup is complemented with programs for conversion, visualization, and analysis of LC-MS/MS-based metabolomics data for metabolomics applications for detection, identification, and quantification of compounds, and statistical analyses [204]. MetaboNexus is an interactive metabolomics data analysis platform that integrates pre-processing of raw peak data with in-depth statistical analysis and metabolite identity searching [205]. MetaboNexus allows users to conduct PCA, PLS-DA, other multi- and univariate analyses (e.g., *t* test, ANOVA, Mann–Whitney *U* test, and Kruskal-Wallis test). In addition it generates graphical outputs, such as score, diagnostic, box, receiver operating characteristic plots, and heat maps, while the metabolite search function uses three major metabolite repositories, HMDB, MassBank and METLIN, using a comprehensive range of molecular adducts. MetaboLyzer is a statistical analysis workflow that handles post-processed LC-MS-based metabolomic data sets and performs a wide range of statistical tests, procedures, and methodologies, in addition to rapid feature identification and biologically relevant analysis via incorporation of four major small molecule databases: KEGG, HMDB, Lipid Maps, and BioCyc [206]. In addition, MetaboLyzer generates heat maps, volcano plots, 3D visualization plots, correlation maps, and metabolic pathway hit histograms.

Developed and maintained by the French Bioinformatics Institute (IFB) and the French Metabolomics and Fluxomics Infrastructure (MetaboHUB), Workflow4Metabolomics (W4M) helps users develop complete workflows for data pre-processing, statistical analysis, and annotation [207]. W4M is an open-source and collaborative online platform for computational metabolomics that is built upon Galaxy and is downloadable as a virtual machine for local installation. Haystack is another online tool for rapid processing and analysis of LC-MS-based metabolomics data and is designed to visualize, parse, filter, and extract significant features [208]. In addition, Haystack predicts class assignment based on PCA and cluster analysis and identifies discriminatory features based on analysis of extracted ion chromatograms (EICs) of binned mass data.

ALLocator is a web-platform for the analysis of LC-ESI-MS experiments which covers the workflow from raw data processing to metabolite identification [209]. This web-platform is especially useful for stable isotopic labeling (SIL) experiments and datasets in which analytes are mostly represented by several adducts and neutral losses. The integrated processing pipeline for spectra deconvolution, ALLocatorSD, generates pseudo-spectra and, in SIL experiments, can

automatically identify peaks emerging from U-^{13}C-labeled internal standard. ALLocator is unique in its variety of interactive functions for manual curation of peak annotations that go hand in hand with automated tools. It is lacking statistical analysis tools by itself, but provides exporting functions for raw and annotated peak tables. MeKO database and software suite provides a platform for evaluation of whether a mutation affects metabolism during normal plant growth and contains images of mutants from the GC-MS data sets, data on differences in metabolite accumulation, and a host of interactive statistical analysis tools (Figure 5) [210]. MassCascade is an open-source library for LC-MS data processing where the available functions have been encapsulated in a plug-in for the workflow environment Konstanz Information Miner (KNIME), allowing combined use with others statistical workflow nodes [211]. The tools discussed above are listed in Table 1.

10. Limitations of current tools and future directions for metabolomics software

The metabolomics research community witnessed the upcoming of a plethora of tools and resources, which the authors do applaud; however, it comes with a price. Currently, many tools are not widely used, and several initiatives by different groups tackle the same challenges described above. This also related to harmonization and standardization of metabolomics experiments and data analysis. For example, if different tools would be able to 'communicate with each other', i.e., have a modular structure where the input and output is in standardized formats, then researchers would be able to build their optimal workflow from the available tools. Another point of attention is the different pre-knowledge and dependencies all these tools require. Here below, we describe some of these aspects in more detail.

The metabolomics tools described in this review were written in different coding languages, each with their own advantages and disadvantages, and using different packages, for example to plot the data on the computer screen. As a result, not all tools are compatible with all operating systems – e.g., some can only be used on Windows operated computers. Furthermore, not all tools will easily 'talk' to each other; though working interfaces are upcoming. In Table 1, we attempted to note the computational dependencies of the metabolomics tools to make the scientist aware of the underlying frameworks these tools use.

A vast majority of the metabolomics tools described above are either written in C, C++, Java, R, Python and MATLAB, while others are far more user friendly with MS Windows OS

compatibility, access from a web server, or by standalone interfaces. For a given programming problem, "scripting languages" (i.e., Perl, Python) can be more productive than conventional languages, and for run time and memory consumption, they are better than Java and C/ C++ [212]. In general, the differences between languages tend to be smaller than the typical differences due to different programmers (individuals) within the same language, although no unambiguous differences in program reliability between the language groups were observed [213]. There is inherent difficulty in integrating MATLAB programs with other software, although tools do exist to address these issues in the system biology context [214]. The comparison of six commonly used programming languages for bioinformatics indicated that a developer should choose an appropriate language carefully, taking into account the performance expected and the library availability for each language [215]. The R Project for Statistical Computing is a freely available software environment [216] that provides a variety of multivariate data analysis and visualization methods and packages. Despite its power, the command-line interface of R is a barrier to its broader use [217]. Additionally, user friendliness, dynamic graphical user interfaces, and less-demanding programming skills would enable 'any' beginner in metabolomics to find the tools immensely useful. For instance, the user friendliness and popularity of Cytoscape, BioGPS, and DAVID for data visualization, integration, and functional enrichment, and the usefulness of Taverna, Kepler, GenePattern, and Galaxy as open-access workbenches for bioinformatics workflows are well known [218]. Undoubtedly, the majority of the tools described in this review are R-programming based (with MATLAB and Java-based tools following-up), implying the growing popularity of the package over Pearl, Python and C++. They also have GUIs which work on the convenient Windows-based platforms, or are simply web-based with user friendliness.

Tools come with heterogeneous features, differential input-output behavior, and variable user interfaces for the majority of tools and resources which still is the major bottleneck for experts and others. To this end, common interfaces, integrated tools, and workflow-based solutions are gaining interest in the community. We expect that open-source software, Cloud-based tools, frameworks, and libraries will become fundamental to the robust analysis and interpretation of metabolomics data, and that the lack of thorough documentation of some tools will result in limited reuse of the software [4]. Large-scale and confident identifications of metabolites, supporting large computational infrastructures, reliable metabolomics databases,

increasing data accessibility and data comparability, integration with other omics, improving sensitivity, dynamic range, and depth-of-coverage, reducing costs, standardization of metabolomics workflows, large scale flux analyses, and metabolite interactions are pertinent issues that have a telling effect on current status of metabolomics research; which is clearly visible by the development of numerous tools to tackle these challenges. In addition, issues that are beyond the scope of this review, such as analytical platform used, metabolite abundances, chromatography, samples, and matrix effects, are reasons which can make great tools behave poorly and unexplainably. One way to create consensus among different laboratories is conducting inter-laboratory studies, for example resulting in 'retention projection' approaches in both LC-MS and GC-MS studies across multiple laboratories [219, 220].

Easier and smoother public sharing of metabolomics data sets, generated from studies ranging from rare types of cancers to near-extinct human populations to hazardous specimens and sensitive epidemiological studies, is needed among other users and collaborators; public domain remains a critical challenge, as discussed at several platforms. Resolving the need for huge repositories across nations, funding agencies, locations, laboratories, and PIs are some of the current challenges limiting better data sharing and collective efforts in annotation, archiving, and biological interpretation. There is an ever increasing demand for high-performance bioinformatics solutions that can help to address the various data processing and data interpretation challenges in metabolomics. And challenges remain in our ability to discern true hits from noise. However, preceding the development of more tools and resources, informaticians and computational biologists need to consider the ease of use and an uncomplicated user interface, while the analytical chemists and mass-spectrometrists/ spectroscopists need to validate the true robustness and applicability of these tools before they can be widely accepted by the metabolomics community and gain popularity. The critical needs of the hour are to reduce the data processing time from hours to minutes (if not seconds), facilitate creation of robust annotation tools and databases, identify the unknown compounds across multiple platforms, and develop more multifunctional tools which would allow a user to start with raw data and end up with biologically meaningful results. Most tools described in this review can give us a good idea of the molecular species one is dealing with, and they are often successful in identifying known and common metabolites. However, the problem with all these approaches is the difficulty or inability to identify the 'unknown unknowns'. Many of the

dedicated databases are limited to a few hundreds of metabolites, whereas there are potentially thousands of metabolites in samples of biological origin, and different packages often lead to identification of different metabolites being identified from the same sets of samples. Many tools have multiple settings that are not properly reported or hardly explained, so that depending on the settings, one can get a vastly different set of metabolites even within the same software package for the same data (i.e., user bias). The need is for identification of metabolites in an unbiased and standardized manner, and the software packages, particularly those that are not commercial, need to be better documented and tested. In an attempt to address this issue, quantitative and alphanumeric metabolite identification metrics were recently introduced [221] to achieve higher confidence with metabolite identification approaches [222]. Additionally, although commercial pipelines are accumulating and becoming available to the global metabolomics research community, for resource-limited laboratories, inexpensive freeware is crucial to progress wide-scale metabolomics research without monetary hindrances. It is expected, then, that community-driven, open-source projects in MS will be the future of metabolomics resource development. It remains to be seen if the above discussed tools will be able to meet the challenges of processing datasets produced by next-generation, mass spectrometry instruments, such as SWATH-MS, Orbitraps, ion-mobility MS, GC-Orbitraps, and myriads of Quadruples helping achieve greater sensitivity and resolution, and survive the test of time. One the one hand, this myriad of tools undoubtedly helps scientist to find the most robust tools to analyze metabolomics data; on the other hand, one could argue that standardization of metabolomics workflows to enhance robust biomarker findings is hampered by this 'explosion' of tools. Open-access data and access to the source codes of tools will hopefully help to come to a more modular manner of building metabolomics workflows such that all groups can use the best tools available for their data sets.

Finally, alongside these metabolomics analysis tools, metabolic flux analysis tools have also developed during this period, such as Central Carbon Metabolic Flux Database (CeCaFDB), for documentation, visualization, and comparative analysis of the quantitative flux results of central carbon metabolism in microbes and animal cells [223]. MetDFBA [224] is a tool for time-resolved metabolomics measurements into dynamic flux balance analysis while thermodynamic elementary flux modes (tEFMA) [225] uses the cellular metabolome to avoid the enumeration of thermodynamically infeasible EFMs, helping in analysis of large-scale metabolic

networks. In addition, Omix Light Edition [226] is an integrated software framework with about 30 modules for graphical modeling, interactive model exploration, and data analysis for all aspects of ^{13}C-metabolic flux analysis [227], but were not covered in this treatise for space constraints. However, alongside metabolomics tools, the fluxomics tools are certainly going to play critical role in systems biology understanding of organisms in the future.

11. Conclusions

The vast majority of the tools that have arisen during this period have focused on the handling, processing, annotation, visualization, and statistical analyses of MS-based metabolomics, as compared to NMR-based, metabolomics datasets. From the above enumeration of tools, software, and resources, it is clearly discernible that, among the steps involved in data collection and analysis, the major bottlenecks are in data processing/ handling and metabolite annotations, for which the number of resources and tools available are growing at the fastest pace. Some additional resources in terms of advances in analytical techniques, multi-platforms data integration, data processing, and quantitative and qualitative analysis of metabolites are available [228]. Systems biology approaches provide meaningful insights into the genotype-phenotype relationship, where multi-omics data generated from genome, transcriptome, proteome, metabolome, and mathematical models are expected to integrate and expand our current understanding of organismal behavior and metabolism. Even so, based on just the number of tools reviewed for each category, it can be safely claimed that metabolite 'annotation' is still the major bottle-neck in metabolomics research. As metabolomics is a rapidly changing field, we can expect new tools to replace the recently developed and existing ones rapidly, as well. However, the ones described here are intended to help the metabolomics community to explore the resources and tools available now and use them to address basic to complex metabolomics challenges. Metabolomics continues to flourish [9], and this compilation of resources will hopefully help to keep us on track and up to date with the most recent significant developments in this big data era and help to keep us moving forward.

Acknowledgments

Biswapriya B. Misra acknowledges a Postdoctoral Fellowship availed at his affiliated institution. In addition, he thanks his extremely helpful colleagues and more importantly Dr.

Drew R. Jones, St. Jude Children's Research Hospital, Memphis, TN; Dr. Dmitry Grapov, Creative Data Solutions; Frederik Walter, CeBiTec, Bielefeld University, Bielefeld; Dr. Chaevien S. Clendinen, Postdoctoral Fellow, Department of Chemistry and Biochemistry, Georgia, among many others for editorial assistance and their personal views and experience with these new and valuable tools. We thank Ms. Barbara Phillips for professional language corrections and editing service. The authors acknowledge the excellent comments and criticisms of the anonymous reviewers and editor for improving the quality of this review. The authors apologizes to all the creators and authors of numerous other *in silico* resources developed during this period, which may have been missed inadvertently or could not find a mention due to space constraints only.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Johnson, S. R., Lange, B. M., Open-access metabolomics databases for natural product research: present capabilities and future potential. *Front. Bioeng. Biotechnol.* 2015, *3*.
2. Patti, G. J., Yanes, O., Siuzdak, G., Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Rev. Mol. Cell Biol.* 2012, *13*, 263-269.
3. Sévin, D. C., Kuehne, A., Zamboni, N., Sauer, U., Biological insights through nontargeted metabolomics. *Curr. Opin. Biotechnol.* 2015, *34*, 1-8.
4. Perez-Riverol, Y., Wang, R., Hermjakob, H., Müller, M., Vesada, V., Vizcaíno, J. A., Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective. *BBA-Proteins Proteom.* 2014, *1844*, 63-76.
5. Alonso, A., Marsal, S., Julià, A., Analytical methods in untargeted metabolomics: state of the art in 2015. *Front. Bioeng. Biotechnol.* 2015, 3.
6. Henry, V. J., Bandrowski, A. E., Pepin, A. S., Gonzalez, B. J., Desfeux, A., OMICtools: an informative directory for multi-omic data analysis. *Database* 2014, bau069.
7. Fiehn Lab Resources, http://fiehnlab.ucdavis.edu/ Accessed on 25th July, 2015

8. Metabolomics Society's Resource Pages, http://metabolomicssociety.org/resources/metabolomics-databases Accessed on 25th July, 2015
9. Roessner, U., Hall, R. D., Metabolomics continues to flourish: highlights from the 2014 Metabolomics Society Conference. *Metabolomics* 2014, *10*, 772-774.
10. Metabomatch, http://www.metabomatch.com/softwares/ Accessed n 25th July, 2015
11. Fuhrer, T., Zamboni, N., High-throughput discovery metabolomics. *Curr. Opin. Biotechnol.* 2015, *31*, 73–78.
12. Codrea, M. C., Jiménez, C. R., Heringa, J., Marchiori, E., Tools for computational processing of LC–MS datasets: a user's perspective. *Comput. Methods Programs Biomed.* 2007, *86*, 281-290.
13. French, W. R., Zimmerman, L. J., Schilling, B., Gibson, B. W., Miller, C. A., Townsend, R. R., Sherrod, S.D., Goodwin, C.R., McLean, J.A., Tabb, D. L. (2014). Wavelet-based peak detection and a new charge inference procedure for MS/MS implemented in ProteoWizard's msConvert. *J. Proteome Res.* 014, *14*, 1299-1307.
14. Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 2010, *11*, 395.
15. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 2006, *78*, 779-787.
16. Tengstrand, E., Lindberg, J., Åberg, K. M., TracMass 2 a modular suite of tools for processing chromatography-full scan mass spectrometry data. *Anal. Chem.* 2014, *86*, 3435-3442.
17. Edmands, W.M., Barupal, D.K., Scalbert, A., MetMSLine: an automated and fully integrated pipeline for rapid processing of high-resolution LC-MS metabolomic datasets. *Bioinformatics*, 2014, *btu705*.
18. Yu, T., Jones, D. P. Improving peak detection in high-resolution LC/MS metabolomics data using preexisting knowledge and machine learning approach. *Bioinformatics* 2014, *30*, 2941-2948.

19. Coble, J. B., Fraga, C. G., Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *J. Chromatogr. A* 2014, *1358*, 155-164.
20. Tsugawa, H., Ohta, E., Izumi, Y., Ogiwara, A., Yukihira, D., Bamba, T., Fukusaki, E., Arita, M. (2014). MRM-DIFF: Data processing strategy for differential analysis in large scale MRM-based lipidomics studies. *Front. Genet.* 2014, *5*, 471.
21. Tsugawa, H., Kanazawa, M., Ogiwara, A., Arita, M., MRMPROBS suite for metabolomics using large-scale MRM assays. *Bioinformatics* 2014, *30*, 2379-2380.
22. Luo, P., Dai, W., Yin, P., Zeng, Z., Kong, H., Zhou, L., Wang, X., Chen, S., Lu, X., Xu, G., MRM-Ion Pair Finder: a systematic approach to transform non-targeted mode to pseudo-targeted mode for metabolomics study based on liquid chromatography-mass spectrometry. *Anal. Chem.* 2015.
23. Nikolskiy, I., Siuzdak, G., Patti, G.J., Discriminating precursors of common fragments for large-scale metabolite profiling by triple quadrupole mass spectrometry. *Bioinformatics*, 2015, *btv085*.
24. Bradbury, J., Genta-Jouve, G., Allwood, J. W., Dunn, W. B., Goodacre, R., Knowles, J. D., He, S., Viant M. R., MUSCLE: automated multi-objective evolutionary optimisation of targeted LC-MS/MS analysis. *Bioinformatics* 2014, btu740.
25. Zhou, R., Tseng, C. L., Huan, T., Li, L., IsoMS: automated processing of LC-MS data generated by a chemical isotope labeling metabolomics platform. *Anal. Chem.* 2014, *86*, 4675-4679.
26. Huan, T., Li, L., Counting missing values in a metabolite-intensity dataset for measuring the analytical performance of a metabolomics platform. *Anal. Chem.* 2014.
27. Huan, T., Li, L., Quantitative metabolome analysis based on chromatographic peak reconstruction in chemical isotope labeling liquid chromatography mass spectrometry. *Anal. Chem.* 2015. doi: : 10.1021/acs.analchem.5b01434
28. Cho, K., Mahieu, N., Ivanisevic, J., Uritboonthai, W., Chen, Y. J., Siuzdak, G., Patti, G. J. isoMETLIN: A database for isotope-based metabolomics. *Anal. Chem.* 2014, *86*, 9358-9361.
29. Neumann, N. K., Lehner, S. M., Kluger, B., Bueschl, C., Sedelmaier, K., Lemmens, M., Krska, R., Schuhmacher, R., Automated LC-HRMS (/MS) approach for the annotation of

fragment ions derived from stable isotope labeling-assisted untargeted metabolomics. *Anal. Chem.* 2014 *86*, 7320-7327.

30. Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., Neumann, S., Trausinger, G., Sinner, F., Pieber T., Magnes, C., IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics* 2015, *16*, 118.
31. Conley, C.J., Smith, R., Torgrip, R.J., Taylor, R.M., Tautenhahn, R., Prince, J.T. Massifquant: open-source Kalman filter-based XC-MS isotope trace feature detection. *Bioinformatics*, 2014, *30*, 2636-2643.
32. Yang, J., Zhao, X., Lu, X., Lin, X., Xu, G., A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. *Front. Mol. Biosci.* 2015, *2*.
33. Nodzenski, M., Muehlbauer, M. J., Bain, J. R., Reisetter, A. C., Lowe, W. L., Scholtens, D. M., Metabomxtr: an R package for mixture-model analysis of non-targeted metabolomics data. *Bioinformatics* 2014, *30*, 3287-3288.
34. http://works.bepress.com/debashis_ghosh/60/ Accessed on 25th July, 2014
35. Zhan, X., Patterson, A. D., Ghosh, D., Kernel approaches for differential expression analysis of mass spectrometry-based metabolomics data. *BMC Bioinformatics* 2015, *16*, 77.
36. Shen, H., Dührkop, K., Böcker, S., Rousu, J. Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics*, 2014, *30*, i157-i164.
37. Fernández-Albert, F., Llorach, R., Garcia-Aloy, M., Ziyatdinov, A., Andres-Lacueva, C., Perera, A., Intensity drift removal in LC/MS metabolomics by common variance compensation. *Bioinformatics* 2014, *30*, 2899-2905.
38. http://github.com/joewandy/peak-grouping-alignment Accessed on 25th July, 2014
39. Wandy, J., Daly, R., Breitling, R., Rogers, S., Incorporating peak grouping information for alignment of multiple liquid chromatography-mass spectrometry datasets. *Bioinformatics* 2015, btv072.
40. Ahmed, S., Zhang, M., Peng, L., GPMS: A genetic programming based approach to multiple alignment of liquid chromatography-mass spectrometry data. In *Applications of Evolutionary Computation*, 2014, 915-927. Springer Berlin Heidelberg.
41. Wehrens, R., Carvalho, E., Fraser, P. D., Metabolite profiling in LC–DAD using multivariate curve resolution: the alsace package for R. *Metabolomics* 2014, *11*, 143-154.

42. Wehrens, R., Bloemberg, T., Eilers, P., Fast parametric time warping of peak lists. *Bioinformatics*, 2015, *btv299*.
43. Suvitaival, T., Rogers, S., Kaski, S., Stronger findings for metabolomics through Bayesian modeling of multiple peaks and compound correlations. *Bioinformatics*, 2014, *30*, i461-i467.
44. Saccenti, E., Hoefsloot, H. C., Smilde, A. K., Westerhuis, J. A., Hendriks, M. M., Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 2014, *10*, 361-374.
45. Xi, B., Gu, H., Baniasadi, H., Raftery, D., Statistical analysis and modeling of mass spectrometry-based metabolomics data. In *Mass Spectrometry in Metabolomics* 2014, 333-353. Springer New York.
46. Triba, M. N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., Rutledge, D. N., Savarin, P. (2015). PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* 2015, *11*, 13-19.
47. Grapov, D., (2014) DeviumWeb: Dynamic Multivariate Data Analysis and Visualization Platform 2014, 10.5281/zenodo.17459, https://github.com/dgrapov/DeviumWeb Accessed on 25th July, 2014
48. Livera, A. M. D., Sysi-Aho, M., Jacob, L., Gagnon-Bartsch, J. A., Castillo, S., Simpson, J. A., Speed, T. P., Statistical methods for handling unwanted variation in metabolomics data. *Anal. Chem.* 2015.
49. Fages, A., Ferrari, P., Monni, S., Dossus, L., Floegel, A., Mode, N., Johansson, M., Travis, R. C., Bamia, C., Sánchez-Pérez, M., Chiodini, P., Boshuizen, H. C., Chadeau-Hyam, M., Riboli, E., Jenab, M., Elena-Herrmann, B., Investigating sources of variability in metabolomic data in the EPIC study: the Principal Component Partial R-square (PC-PR2) method. *Metabolomics* 2014, *10*, 1074-1083.
50. Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E., Edwards, L. M., Metabolomics data normalization with EigenMS. *PloS One* 2014, *9*, e116221.
51. Glaab, E., Schneider, R., RepExplore: Addressing technical replicate variance in proteomics and metabolomics data analysis. *Bioinformatics* 2015, btv127.

52. Chawade, A., Alexandersson, E., Levander, F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* 2014, *13*, 3114-3120.
53. Hughes, G., Cruickshank-Quinn, C., Reisdorph, R., Lutz, S., Petrache, I., Reisdorph, N., Bowler, R., Kechris, K., MSPrep—Summarization, normalization and diagnostics for processing of mass spectrometry–based metabolomic data. *Bioinformatics*, 2014, *30*, 133-134.
54. Jansen, J. J., Blanchet, L., Buydens, L. M., Bertrand, S., Wolfender, J. L., Projected Orthogonalized CHemical Encounter MONitoring (POCHEMON) for microbial interactions in co-culture. *Metabolomics* 2014, 1-12.
55. Cicek, A. E., Roeder, K., Ozsoyoglu, G. MIRA: mutual information-based reporter algorithm for metabolic networks. *Bioinformatics*, 2014, *30,* i175-i184.
56. MassBank, http://www.massbank.jp Accessed on 25th July, 2014
57. Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 2010, *45*, 703-714.
58. METLIN, http://metlin.scripps.edu/ Accessed on 25th July, 2014
59. Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D.E., Abagyan, R., Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug. Monit.* 2005, *27*, 747-751.
60. NIST MS/MS, http://www.nist.gov/mml/bmd/data/tandemmass-speclib.cfm, Accessed on 25th July, 2014
61. Hamdalla, M. A., Mandoiu, I. I., Hill, D. W., Rajasekaran, S., Grant, D. F., BioSM: metabolomics tool for identifying endogenous mammalian biochemical structures in chemical structure space. *J Chem. Inf. Model.* 2013, *53*, 601-612.
62. Hamdalla, M. A., Ammar, R. A., Rajasekaran, S., A molecular structure matching approach to efficient identification of endogenous mammalian biochemical structures. *BMC Bioinformatics* 2015, *16*, S11.

63. Allen, F., Greiner, R., Wishart, D., Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* 2014, 1-13.

64. Allen, F., Pon, A., Wilson, M., Greiner, R., Wishart, D., CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* 2014, *42*, W94-W99.

65. Ogura, T., Bamba, T., Tai, A., Fukusaki, E., Method for the compound annotation of conjugates in nontargeted metabolomics using accurate mass spectrometry, multistage product ion spectra and compound database searching. *Mass Spectrom.* 2015, *4*, A0036-A0036.

66. Lee, H. S., Jo, S., Mukherjee, S., Park, S. J., Skolnick, J., Lee, J., Im, W., GS-align for glycan structure alignment and similarity measurement. *Bioinformatics* 2015, btv202.

67. Fischer, D., Panse, C., Laczko, E. cosmiq: cosmiq - COmbining Single Masses Into Quantities. R package version 1.2.0, http://www.bioconductor.org/packages/devel/bioc/html/cosmiq.html Accessed on 25th July, 2014

68. Mahieu, N. G., Huang, X., Chen, Y. J., Patti, G. J., Credentialing features: A platform to benchmark and optimize untargeted metabolomic methods. *Anal. Chem.* 2014, *86*), 9583-9589.

69. HAMMER, http://www.thermoscientific.com/en/product/mass-frontier-7-0-spectral-interpretation-software.html Accessed on 25th July, 2014

70. Zhou, J., Weber, R. J., Allwood, J. W., Mistrik, R., Zhu, Z., Ji, Z., Chen, S., Dunn, W. B., He, S., Viant, M. R., HAMMER: automated operation of mass frontier to construct in silico mass spectral fragmentation libraries. *Bioinformatics* 2014, *30*, 581-583.

71. Lommen, A., Ultrafast PubChem searching combined with improved filtering rules for elemental composition analysis. *Anal. Chem.* 2014, *86*, 5463-5469.

72. Ahmed, Z., Mayr, M., Zeeshan, S., Dandekar, T., Mueller, M. J., Fekete, A. Lipid-Pro: a computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms. *Bioinformatics* 2014, btu796.

73. Ridder, L., van der Hooft, J. J. J., Verhoeven, S., Automatic compound annotation from mass spectrometry data using MAGMa. *Mass Spectrom.* 2014, *3*, S0033-S0033.

74. Ridder, L., Hooft, J. J. J., Verhoeven, S., Vos, R. C., Schaik, R., Vervoort, J., Substructure-based annotation of high-resolution multistage MS^n spectral trees. *Rapid Commun. Mass Spectrom.* 2012, *26*, 2461-2471.

75. Ridder, L., van der Hooft, J. J. J., Verhoeven, S., de Vos, R. C., Vervoort, J., Bino, R. J. *In silico* prediction and automatic LC–MS n Annotation of green tea metabolites in urine. *Anal. Chem.* 2014, *86*, 4767-4774.

76. Fernández-Albert, F., Llorach, R., Andrés-Lacueva, C., Perera, A. An R package to analyse LC/MS metabolomic data: MAIT (Metabolite Automatic Identification Toolkit). *Bioinformatics*, 2014, *30*, 1937-1939.

77. Dhanasekaran, A. R., Pearson, J. L., Ganesan, B., Weimer, B. C., Metabolome searcher: a high throughput tool for metabolite identification and metabolic pathway mapping directly from mass spectrometry and using genome restriction. *BMC Bioinformatics* 2015, *16*, 62.

78. Zhang, W., Chang, J., Lei, Z., Huhman, D., Sumner, L. W., Zhao, P. X., MET-COFEA: A liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation. *Anal. Chem.* 2014, *86*, 6245-6253.

79. Daly, R., Rogers, S., Wandy, J., Jankevics, A., Burgess, K. E., Breitling, R., MetAssign: probabilistic annotation of metabolites from LC–MS data using a Bayesian clustering approach. *Bioinformatics* 2014, *30*, 2764-2771.

80. Ma, Y., Kind, T., Yang, D., Leon, C., Fiehn, O., MS2Analyzer: A software for small molecule substructure annotations from accurate tandem mass spectra. *Anal. Chem.* 2014, *86*, 10724-10731.

81. Wang, Y., Kora, G., Bowen, B. P., Pan, C., MIDAS: a database-searching algorithm for metabolite identification in metabolomics. *Anal. Chem.* 2014, *86*, 9496-9503.

82. mzCloud database, https://www.mzcloud.org/ Accessed on 25th July, 2014

83. Jeffryes, J., Colestani, R., Elbadawi-Sidhu, MP., Kind, T., Niehaus, T., Broadbelt, L., MS-FINDER: Integrated strategy for structure elucidation on LC-MS/MS by using chemo- and bioinformatics resources. Conference: 11th Annual International Conference of the Metabolomics Society, At San Francisco, California, USA, 2015.

84. Winnikoff, J. R., Glukhov, E., Watrous, J., Dorrestein, P. C., Gerwick, W. H., Quantitative molecular networking to profile marine cyanobacterial metabolomes. *J Antibiot.* 2014, *67*, 105-112.

85. Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M., Moore, B.S., Laskin, J., Bandeira, N., Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* 2012 *109*, E1743-E1752.

86. Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., Wong, W.R., Linington, R.G., Zhang, L., Debonsi, H.M., Gerwick, W.H., Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. *J. Nat. Prod*. 2013, *76*, 1686-1699.

87. Garg, N., Conrad, D., Dorrestein, P., Metabolomics by mass spectrometry based molecular networking and spatial mapping. *FASEB J.* 2015, *29*, 369-371.

88. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods*. 2015.

89. Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Saito, K., Fiehn, O., MS-FINDER: Integrated strategy for structure elucidation on LC-MS/MS by using chemo- and bioinformatics resources. Conference: 11th Annual International Conference of the Metabolomics Society, At San Francisco, California, USA, 2015.

90. Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., Weckwerth, W., mzGroupAnalyzer-Predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PloS One* 2014, *9*, e96188.

91. Silva, R. R., Jourdan, F., Salvanha, D. M., Letisse, F., Jamin, E. L., Guidetti-Gonzalez, S., Labate, C. A., Vêncio, R. Z., ProbMetab: an R package for Bayesian probabilistic annotation of LC–MS-based metabolomics. *Bioinformatics* 2014, *30*, 1336-1337.

92. Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A., Prenni, J. E., RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal. Chem.* 2014, *86*, 6812-6817.

93. Ekanayaka, E. P., Celiz, M. D., Jones, A. D., Relative mass defect filtering of mass spectra: a path to discovery of plant specialized metabolites. *Plant Physiol. 2015*, 167, 1221-1232.

94. Mitchell, J., Fan, T., Lane, A., Moseley, H., Development and *in silico* evaluation of large-scale metabolite identification methods using functional group detection for metabolomics. *The FASEB J.* 2015, *29*, 567-22.

95. Pon, A., Jewison, T., Su, Y., Liang, Y., Knox, C., Maciejewski, A., Wilson, M., Wishart, D. S., Pathways with PathWhiz. *Nucleic Acids Res.* 2015, gkv399.

96. Hamdalla, M. A., Rajasekaran, S., Grant, D. F., Măndoiu, I. I., Metabolic pathway predictions for metabolomics: a molecular structure matching approach. *J. Chem. Inf. Model.* 2015, *55*, 709-718.

97. Kaever, A., Landesfeind, M., Feussner, K., Mosblech, A., Heilmann, I., Morgenstern, B., Feussner, I., Meinicke, P., MarVis-Pathway: integrative and exploratory pathway analysis of non-targeted metabolomics data. *Metabolomics* 2014, 1-14.

98. Eichner, J., Rosenbaum, L., Wrzodek, C., Häring, H. U., Zell, A., Lehmann, R., Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software. *J. Chromatogr. B* 2014, *966*, 77-82.

99. Sun, H., Wang, H., Zhu, R., Tang, K., Gong, Q., Cui, J., Cao, Z., Liu, Q., iPEAP: integrating multiple omics and genetic data for pathway enrichment analysis. *Bioinformatics* 2014, *30*, 737-739.

100. Navas-Delgado, I., García-Godoy, M. J., López-Camacho, E., Rybinski, M., Reyes-Palomares, A., Medina, M. Á., Aldana-Montes, J. F., kpath: integration of metabolic pathway linked data. *Database*, 2015, bav053.

101. Fitzpatrick, M. A., McGrath, C. M., Young, S. P., Pathomx: an interactive workflow-based tool for the analysis of metabolomic data. *BMC Bioinfo.* 2014, *15*, 396.

102. Wachter, A., Beissbarth, T. pwOmics: An R package for pathway-based integration of time-series omics data using public database knowledge. *Bioinformatics*, 2015, btv323.

103. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012, *40*, D109-114.

104. Bolton, E.E., Chen, J., Kim, S., Han, L., He, S., Shi, W., Simonyan, V., Sun, Y., Thiessen, P.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H., PubChem3D: a new resource for scientists. *J. Cheminformatics,* 2011, *3*, 32.

105. Grapov, D., Wanichthanarak, K., Fiehn, O., MetaMapR: pathway independent metabolomic network analysis incorporating unknowns. *Bioinformatics,* 2015, pii: btv194. doi: 10.1093/bioinformatics/btv194

106. Fahrmann, J., Grapov, D., Yang , J., Hammock, B., Fiehn, O., Bell, G.I., Hara, M., Systemic alterations in the metabolome of diabetic NOD mice delineate increased oxidative stress accompanied by reduced inflammation and hypertriglyceridemia. *Am J Physiol Endocrinol. Metab. 308*, E978-89.

107. Wikoff, W.R., Grapov, D., Fahrmann, J.F., DeFelice, B., Rom, W.N., Pass, H.I., Kim, K., Nguyen, U., Taylor, S. L., Gandara, D. R., Kelly, K., Fiehn, O., Miyamoto, S., Metabolomic markers of altered nucleotide metabolism in early stage adenocarcinoma. *Cancer Prev. Res. (Phila.)* 2015, *8*, 410-418.

108. Grapov, D., Fahrmann, J., Hwang, J., Poudel, A., Jo, J., Periwal, V., Fiehn, O., Hara, M., Diabetes associated metabolomic perturbations in NOD mice. *Metabolomics,* 2015, *11*, 425-437.

109. Chemical Translation System, http://cts.fiehnlab.ucdavis.edu/ Accessed on 25th July, 2014

110. Misra, B. B., Acharya, B. R., Granot, D., Assmann, S. M., Chen, S., The guard cell metabolome: functions in stomatal movement and global food security. *Front. Plant Sci.* 2015, *6*, 334.

111. Uppal, K., Soltow, Q. A., Promislow, D. E., Wachtman, L. M., Quyyumi, A. A., Jones, D. P. MetabNet: an R package for metabolic association analysis of high-resolution metabolomics data. *Front Bioeng. Biotechnol.* 2015, *3*, 87.

112. Cytoscape, http://www.cytoscape.org/ Accessed on 25th July, 2014

113. CentiScaPe, http://www.cbmc.it/~scardonig/centiscape/centiscape.php Accessed on 25th July, 2014

114. KEGGscape, http://apps.cytoscape. org/apps/keggscape Accessed on 25th July, 2014

115. GeneMANIA, http://www.genemania.org/ Accessed on 25th July, 2014

116. MetDisease, http://metdisease.ncibi.org Accessed on 25th July, 2014

117. Duren, W., Weymouth, T., Hull, T., Omenn, G. S., Athey, B., Burant, C., Karnovsky, A. MetDisease–connecting metabolites to diseases via literature. *Bioinformatics*, 2014, *btu179*.

118. Fukushima, A., Kanaya, S., Nishida, K., Integrated network analysis and effective tools in plant systems biology. *Front. Plant Sci.* 2014, *5*.

119. Alcaraz, N., Pauling, J., Batra, R., Barbosa, E., Junge, A., Christensen, A. G., Azevedo, V., Ditzel, H.J., Baumbach, J., KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape. *BMC Syst. Biol.* 2014, *8*, 99.

120. Kotera, M., Tabei, Y., Yamanishi, Y., Muto, A., Moriya, Y., Tokimatsu, T., Goto, S., Metabolome-scale prediction of intermediate compounds in multistep metabolic pathways with a recursive supervised approach. *Bioinformatics* 2014, *30*, i165-i174.

121. Narang, P., Khan, S., Hemrom, A. J., Lynn, A. M. MetaNET-a web-accessible interactive platform for biological metabolic network analysis. *BMC Syst. Biol.* 2014, *8*, 130.

122. Pathan, M., Keerthikumar, S., Ang, C. S., Gangoda, L., Quek, C. Y., Williamson, N. A., Mouradov, D., Sieber, O.M., Simpson, R. J., Salim, A., Bacic, A., Hill, A.F., Stroud, D.A., Ryan, M.T., Agbinya, J.I., Mariadasson, J.M., Burgess, A.W., Mathivanan, S. Technical brief funrich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* 2015. doi: 10.1002/pmic.201400515.

123. Turkarslan, S., Wurtmann, E. J., Wu, W. J., Jiang, N., Bare, J. C., Foley, K., Reiss, D.J., Novichkov, P., Baliga, N. S., Network portal: a database for storage, analysis and visualization of biological networks. *Nucleic Acids Res.* 2014, *42*, D184-D190.

124. Yamanishi, Y., Tabei, Y., Kotera, M. Metabolome-scale *de novo* pathway reconstruction using regioisomer-sensitive graph alignments. *Bioinformatics*, 2015, *31*, i161-i170.

125. Pertusi, D.A., Stine, A.E., Broadbelt, L.J., Tyo, K.E., Efficient searching and annotation of metabolic networks using chemical similarity. *Bioinformatics*, 2014, *btu760*.

126. Moreno, P., Beisken, S., Harsha, B., Muthukrishnan, V., Tudose, I., Dekker, A., Dornfeldt, S., Taruttis, F., Grosse, I., Hastings, J., Neumann, S., Steinbeck, C., BiNChE: A web tool and library for chemical enrichment analysis based on the ChEBI ontology. *BMC Bioinformatics* 2015, *16*, 56.

127. Seaver, S. M., Gerdes, S., Frelin, O., Lerma-Ortiz, C., Bradbury, L. M., Zallot, R., Hasnain, G., Niehaus, T. D., El Yacoubi, B., Pasternak, S., Olson, R., Pusch, G., Overbeek, R., Stevens, R., de Crécy-Lagard, V., Ware, D., Hanson, A.D., Henry, C. S., High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. *Proc. Natl. Acad. Sci. U S A* 2014, *111*, 9645-9650.

128. Mak, T. D., Laiakis, E. C., Goudarzi, M., Fornace Jr, A. J., Selective paired ion contrast analysis: A novel algorithm for analyzing postprocessed LC-MS metabolomics data possessing high experimental noise. *Anal. Chem.* 2015, *87*, 3177-3186.

129. Costa, R. S., Veríssimo, A., Vinga, S., KiMoSys: a web-based repository of experimental data for KInetic MOdels of biological SYStems. *BMC Syst. Biol.* 2014, *8*, 85.

130. Carazzolle, M. F., de Carvalho, L. M., Slepicka, H. H., Vidal, R. O., Pereira, G. A. G., Kobarg, J., Meirelles, G. V., IIS–Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein-metabolite-gene-drug interactions by integrating a variety of data sources and tools. *PLoS One* 2014, *9*, e100385.

131. Röst, H. L., Rosenberger, G., Aebersold, R., Malmström, L., Efficient visualization of high-throughput targeted proteomics experiments: TAPIR. *Bioinformatics*, 2015, *btv152*.

132. Marttinen, P., Pirinen, M., Sarin, A. P., Gillberg, J., Kettunen, J., Surakka, I., Kangas, A.J., Soininen, P., O'Reilly, P., Kaakinen, M., Kähönen, M., Lehtimäki, T., Ala-Korpela, M., Raitakari, O.T., Salomaa, V., Järvelin, M.R., Ripatti, S., Kaski, S., Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. *Bioinformatics*, 2014, *btu140*.

133. Bolton, E. E., Wang, Y., Thiessen, P. A., Bryant, S. H., PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* 2008, *4*, 217-241.

134. Pence, H. E., Williams, A., ChemSpider: an online chemical information resource. *J. Chem. Educ.* 2010, *87*, 1123-1124.

135. Zhu, Z. J., Schultz, A. W., Wang, J., Johnson, C. H., Yannone, S. M., Patti, G. J., Siuzdak, G. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat. Prot.* 2013, *8*, 451-460.

136. KNApSAcK Family Databases, http://kanaya.naist.jp/KNApSAcK_Family/

137. Ohtana, Y., Abdullah, A. A., Altaf-Ul-Amin, M., Huang, M., Ono, N., Sato, T., Sugiura, T., Horai, H., Nakamura, Y., Morita (Hirai), A., Lange, K. W., Kibinge, N. K., Katsuragi,

T., Shirai, T., Kanaya, S., Clustering of 3D-structure similarity based network of secondary metabolites reveals their relationships with biological activities. *Mol. Inform.* 2014, *33*, 790-801.

138. Yue, Y., Chu, G. X., Liu, X. S., Tang, X., Wang, W., Liu, G. J., Yang, T., Ling, T. J., Wang, X. G., Zhang, Z. Z., Xia, T., Wan, X. C., Bao, G. H., TMDB: A literature-curated database for small molecular compounds found from tea. *BMC Plant Biol.* 2014, *14*, 243.

139. Sharma, A., Dutta, P., Sharma, M., Rajput, N. K., Dodiya, B., Georrge, J. J., Kholia, T., OSDD Consortium, Bhardwaj, A., BioPhytMol: a drug discovery community resource on anti-mycobacterial phytomolecules and plant extracts. *J. Cheminform.* 2014, *6*, 1-10

140. Kumari, S., Pundhir, S., Priya, P., Jeena, G., Punetha, A., Chawla, K., Firdos Jafaree, Z., Mondal, S., Yadav, G., EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom. *Database* 2014, bau120.

141. Lemfack, M. C., Nickel, J., Dunkel, M., Preissner, R., Piechulla, B., mVOC: a database of microbial volatiles. *Nucleic Acids Res.* 2014, *42*, D744-D748.

142. Viant, M. R., Metabolomics of aquatic organisms: the new'omics' on the block. *Mar. Ecol. Prog. Ser.* 2007, *332*, 301-306.

143. Longnecker, K., Futrelle, J., Coburn, E., Soule, M. C. K., Kujawinski, E. B. (2015). Environmental metabolomics: Databases and tools for data analysis. *Mar. Chem.* 2015. doi:10.1016/j.marchem.2015.06.012

144. Carroll, A. J., Zhang, P., Whitehead, L., Kaines, S., Tcherkez, G., Badger, M. R. PhenoMeter: a metabolome database search tool using statistical similarity matching of metabolic phenotypes for high-confidence detection of functional links. *Front. Bioeng. Biotechnol.*, 2015, *3*, 106.

145. Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S.A., Rappaport, S.M. (2015). T3DB: the toxic exposome database. *Nucleic Acids Res.*, 2015, *43*, D928-D934.

146. Aimo, L., Liechti, R., Nouspikel, N., Niknejad, A., Gleizes, A., Götz, L., Kuznetsov, D., David, F.P., van der Goot, F.G., Riezman, H., Bougueleret, L., Xenarios, I., Bridge, A. The SwissLipids knowledgebase for lipid biology. *Bioinformatics*, 2015, *btv285*.

147. Sneddon, T. P., Zhe, X. S., Edmunds, S. C., Li, P., Goodman, L., Hunter, C. I., GigaDB: promoting data dissemination and reproducibility. *Database*, 2014, bau018.

148. Haug, K., Salek, R. M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P., Maguire, E., González-Beltrán, A., Sansone, S.A., Griffin, J.L., Steinbeck, C. MetaboLights- an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 2012, gks1004.

149. Haug, K., Salek, R. M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P., Maguire, E., González-Beltrán, A., Sansone, S-A., Griffin, J.L., Steinbeck, C., MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 2012, gks1004.

150. Metabolomics Workbench (http://www.metabolomicsworkbench.org/ Accessed on 9th September, 2015.

151. MetabolomeXchange (http://metabolomexchange.org/site/) Accessed n 9th September, 2015.

152. Kirwan, J. A., Weber, R. J., Broadhurst, D. I., Viant, M. R., Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Scientific Data*, 2014, *1*.

153. Sansone, S. A., Rocca-Serra, P., Brandizi, M., Brazma, A., Field, D., Fostel, J., Garrow, A.G., Gilbert, J., Goodsaid, F., Hardy, N., Jones, P., Lister, A., Miller, M., Morrison, N., Rayner, T., Sklyar, N., Taylor, C., Tong, W., Warner, G., Wiemann, S., The first RSBI (ISA-TAB) workshop: "Can a simple format work for complex studies?". *OMICS Journal* 2008, *12*, 143-149.

154. González-Beltrán, A., Maguire, E., Sansone, S. A., Rocca-Serra, P., linkedISA: semantic representation of ISA-Tab experimental metadata. *BMC Bioinformatics* 2014, *15*, S4.

155. González-Beltrán, A., Neumann, S., Maguire, E., Sansone, S. A., Rocca-Serra, P., The Risa R/Bioconductor package: integrative data analysis from experimental metadata and back again. *BMC Bioinformatics* 2014, *15*, S11.

156. Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L.W., Goodacre, R., Hardy, N.W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J.C., Sansone, S.S., The metabolomics standards initiative (MSI). *Metabolomics* 2007, *3*, 175-178.

157. Sansone, S. A., Fan, T., Goodacre, R., Griffin, J. L., Hardy, N. W., Kaddurah-Daouk, R., Kristal, B.S., Lindon, J., Mendes, P., Morrison, N., Nikolau, B., Robertson, D., Sumner, L.W., Taylor, C., van der Werf, M., van Ommen, B., Fiehn, O., The metabolomics standards initiative. *Nat. Biotechnol.* 2007, *25*, 846-848.

158. Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T.W., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M. R. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 2007, *3*, 211-221.

159. Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D. Y., Lu, Y., Moon, S., Nikolau, B., Quality control for plant metabolomics: reporting MSI-compliant studies. *The Plant J.* 2008, *53*, 691-704.

160. Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R., Dunn, W. B., The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience* 2013, *2*, 13-13.

161. Ara, T., Enomoto, M., Arita, M., Ikeda, C., Kera, K., Yamada, M., Nishioka, T., Ikeda, T., Nihei, Y., Shibata, D., Kanaya, S., Sakurai, N., Metabolonote: A wiki-based database for managing hierarchical metadata of metabolome analyses. *Front Bioeng. Biotechnol.* 2015, *3*.

162. Sakurai, N., Ara, T., Enomoto, M., Motegi, T., Morishita, Y., Kurabayashi, A., Iijima, Y., Ogata, Y., Nakajima, D., Suzuki, H., Shibata, D., Tools and databases of the KOMICS web portal for preprocessing, mining, and dissemination of metabolomics data. *Biomed Res. Int.* 2014.

163. Wehrens, R., Weingart, G., Mattivi, F., metaMS: An open-source pipeline for GC–MS-based untargeted metabolomics. *J. Chromatogr. B* 2014, *966*, 109-116.

164. Franceschi, P., Mylonas, R., Shahaf, N., Scholz, M., Arapitsas, P., Masuero, D., Weingart, G., Carlin, S., Vrhovsek, U., Mattivi, F., Wehrens, R., MetaDB a data processing workflow in untargeted MS-based metabolomics experiments. *Front Bioeng. Biotechnol.* 2014, *2*.

165. Palla, P., Frau, G., Vargiu, L., Rodriguez-Tomé, P. (2014). QTREDS: a Ruby on Rails-based platform for omics laboratories. *BMC Bioinformatics*, 2014, *15*, S13.

166. Mastr-MS, https://mastr-ms.readthedocs.org/en/latest/ Accessed on 25th July, 2014

167. Nahid, A., MASTR-MS: Metabolomics Australia Sample Tracking System. In *PAG Asia 2014*. Plant and Animal Genome.

168. Yabi, https://ccg.murdoch.edu.au/yabi/login/?next=/yabi/ Accessed on 25th July, 2014

169. Barrero, R., Yabi: A Cross–Omics Internet-Based Platform for Bioinformatics Workflows. In *PAG Asia* 2014. Plant and Animal Genome.

170. Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 2008, *24*, 2534-2536.

171. Griss, J., Jones, A. R., Sachsenberg, T., Walzer, M., Gatto, L., Hartler, J., Thallinger, G. G., Salek, R. M., Steinbeck, C., Neuhauser, N., Cox, J., Neumann, S., Fan, J., Reisinger, F., Xu, Q. W., Del Toro, N., Pérez-Riverol, Y., Ghali, F., Bandeira, N., Xenarios, I., Kohlbacher, O., Vizcaíno, J. A., Hermjakob, H., The mzTab data exchange format: Communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Mol Cell Proteomics* 2014, *13*, 2765-2775.

172. Xu, Q. W., Griss, J., Wang, R., Jones, A. R., Hermjakob, H., Vizcaíno, J. A., jmzTab: A Java interface to the mzTab data standard. *Proteomics* 2014, *14*, 1328-1332.

173. LipidDataAnalyzer, http://genome.tugraz.at/lda/ Accessed on 25th July, 2014

174. Röst, H. L., Schmitt, U., Aebersold, R., Malmström, L., pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library. *Proteomics*, 2014, *14*, 74-77.

175. Liu, Y., Liang, Y., Wishart, D., PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res.* 2015, gkv383. doi: 10.1093/nar/gkv383

176. Beisken, S., Conesa, P., Haug, K., Salek, R. M., Steinbeck, C., SpeckTackle: JavaScript charts for spectroscopy. *J. Cheminform.* 2015, *7*, 17. doi: 10.1186/s13321-015-0065-7

177. Garcia-Albornoz, M., Thankaswamy-Kosalai, S., Nilsson, A., Väremo, L., Nookaew, I., Nielsen, J., BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data. *Nucleic Acids Res.*, 2014, gku371.

178. Heyman, H.M., Dubery, I.A., The potential of mass spectrometry imaging in plant metabolomics: a review. *Phytochem. Rev.,* 2015, 1-20.

179. Horn, P. J., Chapman, K. D., Metabolite Imager: customized spatial analysis of metabolite distributions in mass spectrometry imaging. *Metabolomics* 2014, *10*, 337-348.

180. Wijetunge, C. D., Saeed, I., Boughton, B. A., Spraggins, J. M., Caprioli, R. M., Bacic, A., Roessner, U., Halgamuge, S.K., EXIMS: an improved data analysis pipeline based on a new peak picking method for EXploring Imaging Mass Spectrometry data. *Bioinformatics*, 2015, *btv356*.

181. Kuich, P., Hoffmann, N., Kempa, S., Maui-VIA: A user-friendly software for visual identification, alignment, correction, and quantification of gas chromatography–mass spectrometry data. *Front. Bioeng. Biotechnol.* 2015, *2*, 84.

182. Domingo-Almenara, X., Perera, A., Ramírez, N., Brezmes, J., Compound identification in comprehensive gas chromatography—mass spectrometry-based metabolomics by blind source separation. In *9th International Conference on Practical Applications of Computational Biology and Bioinformatics* (pp. 49-57). Springer International Publishing. 2015. doi: 10.1007/978-3-319-19776-0_6

183. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., Neumann, S., CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* 2011, *84*, 283-289.

184. Aggio, R. B., Mayor, A., Reade, S., Probert, C. S., Ruggiero, K., Identifying and quantifying metabolites by scoring peaks of GC-MS data. *BMC Bioinformatics* 2014, *15*, 374.

185. Hoffmann, N., Wilhelm, M., Doebbe, A., Niehaus, K., Stoye, J. BiPACE 2D-graph-based multiple alignment for comprehensive 2D gas chromatography-mass spectrometry. *Bioinformatics*, 2014, *30*, 988-995.

186. Dubey, A., Rangarajan, A., Pal, D., Atreya, H. S., A pattern recognition based approach for identifying metabolites in NMR based metabolomics. *Anal. Chem.* 2015. doi: 10.1021/acs.analchem.5b00990

187. Alonso, A., Rodríguez, M. A., Vinaixa, M., Tortosa, R., Correig, X., Julià, A., Marsal, S. Focus: a robust workflow for one-dimensional NMR spectral analysis. *Anal. Chem.,* 2013*, 86*, 1160-1169.

188. Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., Ebbels, T. M., Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc.* 2014, *9*, 1416-1427.

189. Jauhiainen, A., Basetti, M., Narita, M., Narita, M., Griffiths, J., Tavaré, S., Normalization of metabolomics data with applications to correlation maps. *Bioinformatics* 2014, btu175.

190. Ravanbakhsh, S., Liu, P., Bjorndahl, T., Mandal, R., Grant, J., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., Wishart, D., Fully-automated NMR spectral profiling for metabolomics. *PLoS One* 2015, *10*, e0124219

191. Jayaseelan, K. V., Steinbeck, C., Building blocks for automated elucidation of metabolites: natural product-likeness for candidate ranking. *BMC Bioinformatics* 2014, *15*, 234.

192. Bingol, K., Li, D. W., Bruschweiler-Li, L., Cabrera, O. A., Megraw, T., Zhang, F., Brüschweiler, R., Unified and Isomer-specific NMR metabolomics database for the accurate analysis of ^{13}C–^{1}H HSQC spectra. *ACS Chem. Biol.* 2014.

193. Bingol, K., Bruschweiler-Li, L., Li, D. W., Brüschweiler, R., Customized metabolomics database for the analysis of NMR ^{1}H–^{1}H TOCSY and ^{13}C–^{1}H HSQC-TOCSY spectra of complex mixtures. *Anal. Chem.* 2014, *86*, 5494-5501.

194. Hedjazi, L., Gauguier, D., Zalloua, P., Nicholson, J. K., Dumas, M. E., Cazier, J. B., mQTL. NMR: An integrated suite for genetic mapping of quantitative variations of 1H NMR-based metabolic profiles. *Anal. Chem.* 2015.

195. Worley, B., Powers, R., MVAPACK: a complete data handling package for NMR metabolomics. *ACS Chem. Biol.* 2014, *9*, 1138-1144.

196. Hanson, B. A., ChemoSpec: An R package for chemometric analysis of spectroscopic data (Package version 4.0. 1), 2015. Accessed on 25th July, 2014

197. Bingol, K., Bruschweiler, R., NMR/MS Translator for the enhanced simultaneous analysis of metabolomics mixtures by NMR spectroscopy and mass spectrometry: Application to human urine. *J. Proteome. Res.* 2015, *14*, 2642-2648.

198. Clendinen, C. S., Pasquel, C., Ajredini, R., Edison, A. S., ^{13}C NMR metabolomics: INADEQUATE network analysis. *Anal. Chem.* 2015. doi: 10.1021/acs.analchem.5b00867.

199. MetaboAnalyst, version 3.0, www.metaboanalyst.ca

200. Xia, J., Sinelnikov, I. V., Han, B., Wishart, D. S., MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* 2015, gkv380.

201. Gowda, H., Ivanisevic, J., Johnson, C. H., Kurczy, M. E., Benton, H. P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevalo, B., Westenskow, P. D., Wang, J., Arkin, A. P.,

Deutschbauer, A. M., Patti, G. J., Siuzdak, G., Interactive XCMS Online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* 2014, *86*, 6931-6939.

202. Benton, H. P., Ivanisevic, J., Mahieu, N. G., Kurczy, M. E., Johnson, C. H., Franco, L., Rinehart, D., Valentine, E., Gowda, H., Ubhi, B.K., Tautenhahn, R., Gieschen, A., Fields, M.W., Patti, G.J., Siuzdak, G., Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal. Chem.* 2014, 87, 884-891.

203. Tanaka, S., Fujita, Y., Parry, H. E., Yoshizawa, A. C., Morimoto, K., Murase, M., Yamada, Y., Yao, J., Utsunomiya, S.I., Kajihara, S., Fukuda, M., Ikawa, M., Tabata, T., Takahashi, K., Aoshima, K., Nihei, Y., Nishioka, T., Oda, Y., Tanaka, K., Mass++: A visualization and analysis tool for mass spectrometry. *J. Proteome Res.* 2014, *13*, 3846-3853.

204. Winkler, R., MASSyPup—an 'Out of the Box'solution for the analysis of mass spectrometry data. *J. Mass Spectrom.* 2014, *49*, 37-42.

205. Huang, S. M., Toh, W., Benke, P. I., Tan, C. S., Ong, C. N., MetaboNexus: An interactive platform for integrated metabolomics analysis. *Metabolomics* 2014, *10*, 1084-1093.

206. Mak, T. D., Laiakis, E. C., Goudarzi, M., Fornace Jr, A. J., Metabolyzer: a novel statistical workflow for analyzing postprocessed LC–MS metabolomics data. *Anal. Chem.* 2014, *86*, 506-513.

207. Giacomoni, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., Duperier, C., Tremblay-Franco, M., Martin, J. F., Jacob, D., Goulitquer, S., Thévenot, E. A., Caron, C., Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics* 2014, btu813.

208. Grace, S. C., Embry, S., Luo, H., Haystack, a web-based tool for metabolomics research. *BMC Bioinformatics* 2014, *15*, S12.

209. Kessler, N., Walter, F., Persicke, M., Albaum, S. P., Kalinowski, J., Goesmann, A., Niehaus, K., Nattkemper, T. W., ALLocator: An interactive web platform for the analysis of metabolomic LC-ESI-MS datasets, enabling semi-automated, user-revised compound annotation and mass isotopomer ratio analysis. *PloS One* 2014, *9*, e113909.

210. Fukushima, A., Kusano, M., Mejia, R. F., Iwasa, M., Kobayashi, M., Hayashi, N., Watanabe-Takahashi, A., Narisawa, T., Tohge, T., Hur, M., Wurtele, E. S., Nikolau, B. J., Saito, K., Metabolomic characterization of knockout mutants in *Arabidopsis*: development

of a metabolite profiling database for knockout mutants in *Arabidopsis*. *Plant Physiol.* 2014, *165*, 948-961.

211. Beisken, S., Earll, M., Portwood, D., Seymour, M., Steinbeck, C., MassCascade: Visual Programming for LC-MS Data Processing in Metabolomics. *Mol. Inform.* 2014, *33*, 307-310.

212. Prechelt, L., Are scripting languages any good? A validation of Perl, Python, Rexx, and Tcl against C, C++, and Java. *Adv. Comput.* 2003, *57*, 205-270.

213. Prechelt, L., An empirical comparison of C, C++, Java, Perl, Python, Rexx and Tcl. *IEEE Computer*. 2000, *33*, 23-29.

214. Wellock, C., Chickarmane, V., Sauro, H. M., The SBW–MATLAB interface. *Bioinformatics* 2005, *21*, 823-824.

215. Fourment, M., Gillings, M. R., A comparison of common programming languages used in bioinformatics. *BMC Bioinformatics* 2008, *9*, 82.

216. Team, R. C. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2014.

217. Grapov, D., Newman, J. W., imDEV: a graphical user interface to R multivariate analysis tools in Microsoft Excel. *Bioinformatics* 2012, *28*, 2288-2290.

218. Kouskoumvekaki, I., Shublaq, N., Brunak, S., Facilitating the use of large-scale biological data and tools in the era of translational bioinformatics. *Brief. Bioinform.* 2014, *15*, 942-952.

219. Abate-Pella, D., Freund, D. M., Ma, Y., Simón-Manso, Y., Hollender, J., Broeckling, C. D., Huhman, D. V., Krokhin, O. V., Stoll, D. R., Hegeman, A. D., Kind, T., Fiehn, O., Schymanski, E. L., Prenni, J. E., Sumner, L. W., Boswell, P. G. Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. *Journal of Chromatography A* 2015.

220. Barnes, B. B., Wilson, M. B., Carr, P. W., Vitha, M. F., Broeckling, C. D., Heuberger, A. L., Prenni, J., Janis, G. C., Corcoran, H., Snow, N.H., Chopra, S., Dhandapani, R., Tawfall, A., Sumner, L. W., Boswell, P. G., Boswell, P. G. "Retention projection" enables reliable use of shared gas chromatographic retention data across laboratories, instruments, and methods. *Analytical chemistry*, 2013, *85*, 11650-11657.

221. Sumner, L. W., Lei, Z., Nikolau, B. J., Saito, K., Roessner, U., Trengove, R. Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics* 2014, *10*, 1047-1049.

222. Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E.L., Sumner, L.W., Tengrove, R., Wolfender, J. L., Metabolite identification: are you sure? And how do your peers gauge your confidence?. *Metabolomics* 2014, *10*, 350-353.

223. Zhang, Z., Shen, T., Rui, B., Zhou, W., Zhou, X., Shang, C., Xin, C., Liu, X., Li, G., Jiang, J., Li, C., Li, R., Han, M., You, S., Yu, G., Yi, Y., Wen, H., Liu, Z., Xie, X., CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by ^{13}C-fluxomics. *Nucleic Acids Res.*, 2014, *gku1137*.

224. Willemsen, A. M., Hendrickx, D. M., Hoefsloot, H. C., Hendriks, M. M., Wahl, S. A., Teusink, B., Smilde, A. K., van Kampen, A. H., MetDFBA: incorporating time-resolved metabolomics measurements into dynamic flux balance analysis. *Mol. BioSyst.* 2015, *11*, 137-145.

225. Gerstl, M. P., Jungreuthmayer, C., Zanghellini, J. tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*, 2015, *btv111*.

226. Omix Light Edition, http://www.omix-visualization.com/#sthash.SlwVYbuJ.dpbs

227. Nöh, K., Droste, P., Wiechert, W., Visual workflows for 13C-metabolic flux analysis. *Bioinformatics*, 2015, *31*, 346-354.

228. Duan, L. X., Qi, X., Metabolite qualitative methods and the introduction of metabolomics database. In *Plant Metabolomics* 2015, 171-193. Springer Netherlands.

Figures Legends:

Figure 1. A. Typical output for OPLS-DA at DeviumWeb statistical interface; and **B.** MetaMapR aided mapping of structural (Tanimoto - yellow edges) and biochemical (KEGG - blue edges) functions of the metabolites helps the visualization of metabolite-metabolite networks.

Figure 2. A. RepExplore analysis of metabolomics data sets from *A. thaliana* mutant and wild type showing bar plots; and **B.** a heat map of the same study; and **C.** Typical output obtained using the Normalyzer tool for metabolomics data sets using 100 samples (n=4 biological replicates) each with 300 metabolites and all available normalization methods.

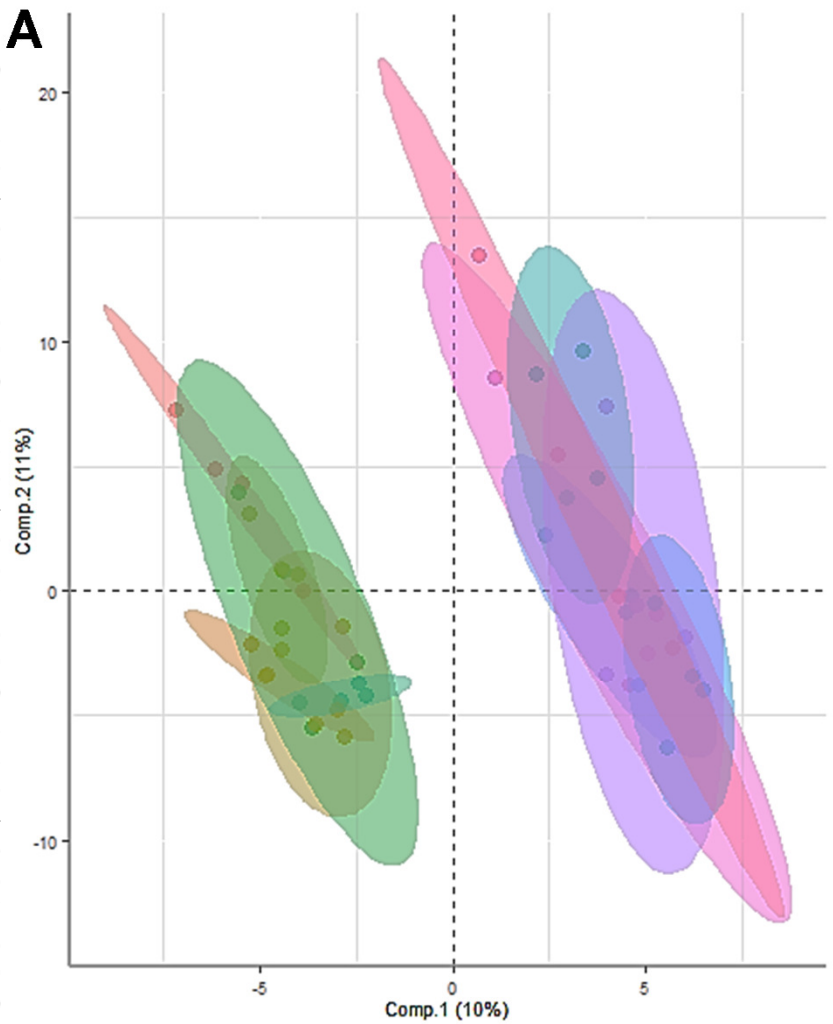
Figure 3. The MzCloud tool for MS/MS data annotation. The example shows the MS1 spectrum for 'mannitol' and its spectral tree down to the MS4 level.

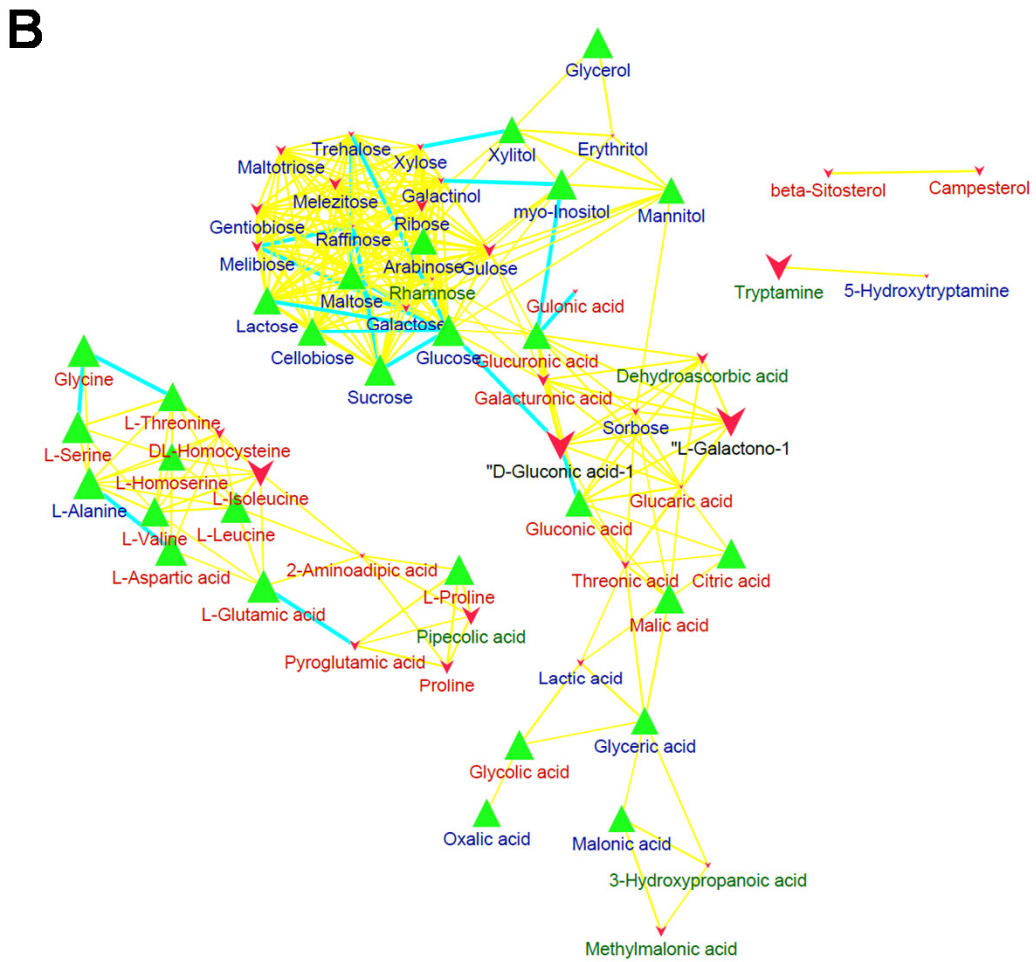
Figure 4. Bayesil analysis for NMR-based metabolomics, with: **A.** Spectra viewer displaying the processed spectrum (black) and the fit (blue) from filtered serum samples obtained using a 600 MHz Bruker instrument (this corresponds to 'Example 3' provided on the Bayesil website); **B.** Downloadable .csv file provides the HMDB IDs, compound concentrations, threshold, and confidence score leading to annotation of metabolomics samples; and **C**. Bayesil user interface showing the parameters an user needs to submit for analyses of NMR data.

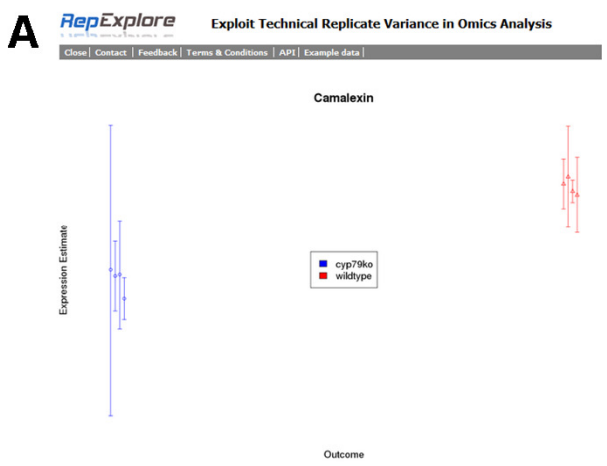
Figure 5. MeKO tool at RIKEN PRIMe web server provides tools for statistical and visual interpretation of metabolomics datasets. Shown is the heat map clustering of metabolomic data for *Arabidopsis thaliana* obtained from a gas chromatography-time-of-flight/mass spectrometry (GC-TOF/MS).

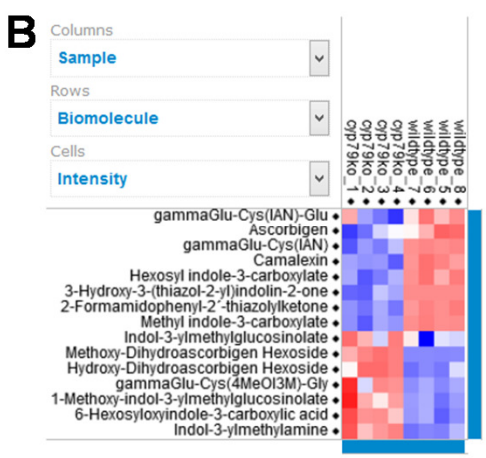
Table Legends:

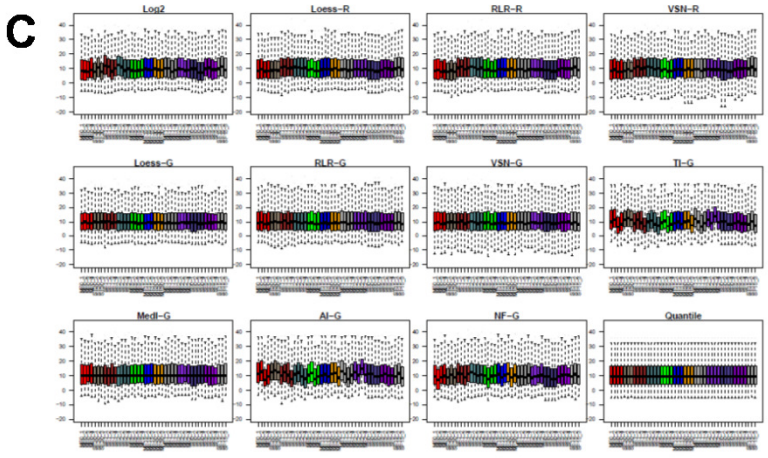
Table 1. Table enlisting the metabolomics tools and resources developed during the period 2014-15, consisting of the name, and displaying their platform dependencies in terms of analytical input and computational dependencies, web address (URL), and their reference if available. Abbreviations used: Any: input data from different kinds of analytical platforms, GC: gas chromatography, Internet: tool running on server and assessable by online access, LC: liquid chromatography, MS: mass spectrometry, MS/MS: mass spectrometry fragmentation, NMR: nuclear magnetic resonance spectroscopy, R: R package, UV-DAD: ultraviolet diode array detection, - : not applicable/available.

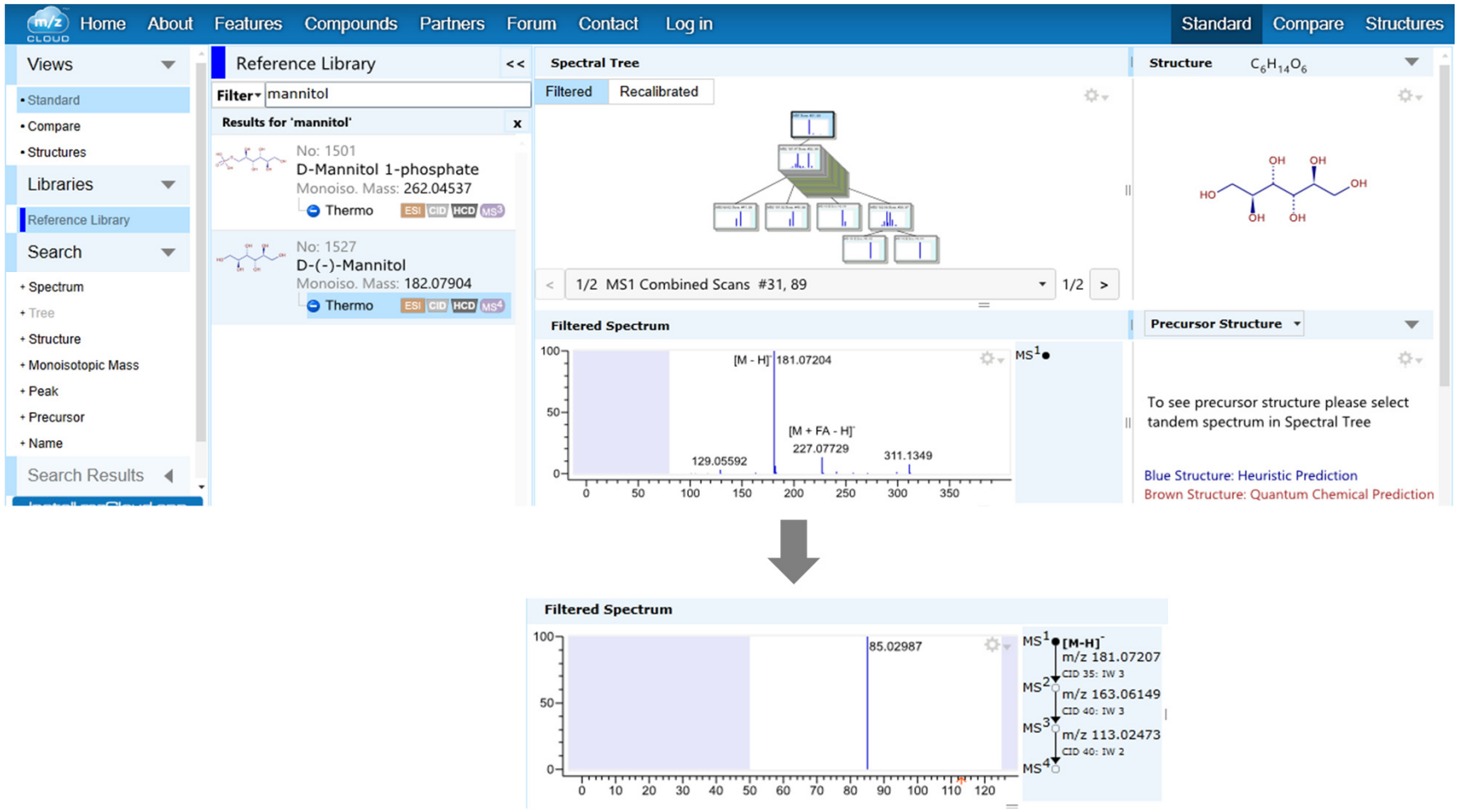

A
20
10
0
-10
Comp.2 (11%)
-5
0
5
Comp.1 (10%)


B
Glycerol
Trehalose
Xylose
Xylitol
Erythritol
Maltotriose
Melezitose
Galactinol
myo-Inositol
Mannitol
Gentiobiose
Raffinose
Ribose
Melibiose
Arabinose
Gulose
Maltose
Rhamnose
Gulonic acid
Lactose
Galactose
Cellobiose
Glucose
Glucuronic acid
Sucrose
Dehydroascorbic acid
Galacturonic acid
Sorbose
"L-Galactono-1
"D-Gluconic acid-1
Glucaric acid
Gluconic acid
Threonic acid
Citric acid
Malic acid
Lactic acid
Glyceric acid
Glycolic acid
Oxalic acid
Malonic acid
3-Hydroxypropanoic acid
Methylmalonic acid
beta-Sitosterol
Campesterol
Tryptamine
5-Hydroxytryptamine
Glycine
L-Threonine
L-Serine
DL-Homocysteine
L-Homoserine
L-Isoleucine
L-Alanine
L-Valine
L-Leucine
L-Aspartic acid
2-Aminoadipic acid
L-Proline
L-Glutamic acid
Pipecolic acid
Pyroglutamic acid
Proline


A
RepExplore
Exploit Technical Replicate Variance in Omics Analysis
Close | Contact | Feedback | Terms & Conditions | API | Example data |
Camalexin
Expression Estimate
cyp79ko
wildtype
Outcome

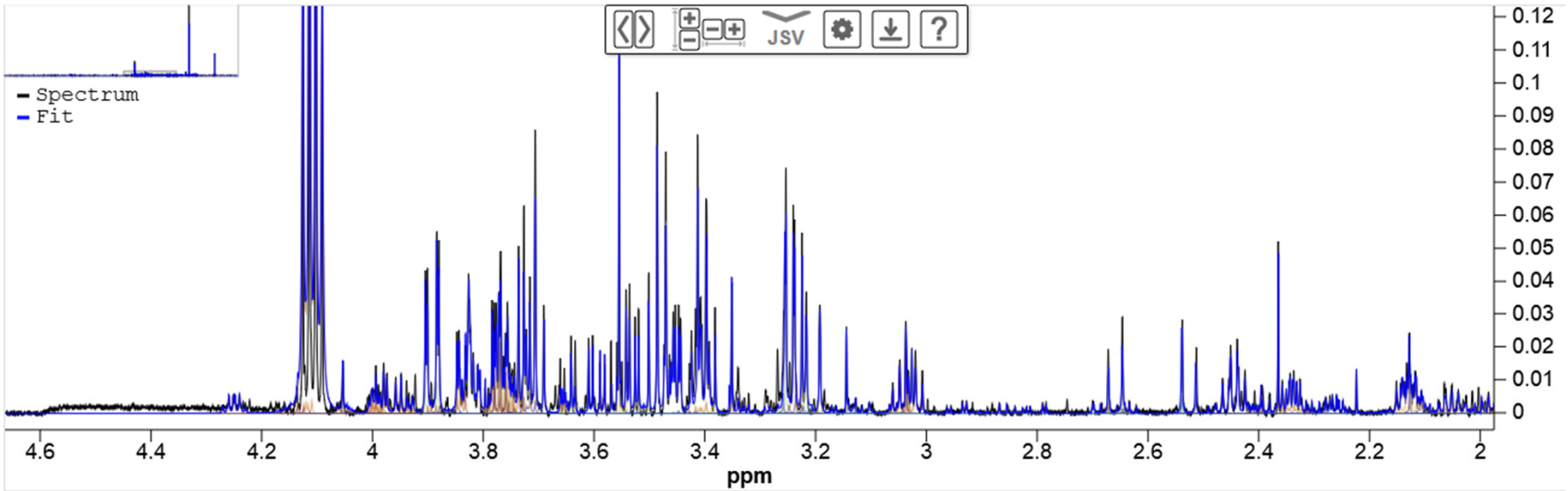

B
Columns
Sample
Rows
Biomolecule
Cells
Intensity
cyp79ko_1
cyp79ko_2
cyp79ko_3
cyp79ko_4
wildtype_7
wildtype_6
wildtype_5
wildtype_8
gammaGlu-Cys(IAN)-Glu
Ascorbigen
gammaGlu-Cys(IAN)
Camalexin
Hexosyl indole-3-carboxylate
3-Hydroxy-3-(thiazol-2-yl)indolin-2-one
2-Formamidophenyl-2´-thiazolylketone
Methyl indole-3-carboxylate
Indol-3-ylmethylglucosinolate
Methoxy-Dihydroascorbigen Hexoside
Hydroxy-Dihydroascorbigen Hexoside
gammaGlu-Cys(4MeOI3M)-Gly
1-Methoxy-indol-3-ylmethylglucosinolate
6-Hexosyloxyindole-3-carboxylic acid
Indol-3-ylmethylamine


C
Log2
Loess-R
RLR-R
VSN-R
Loess-G
RLR-G
VSN-G
TI-G
MedI-G
AI-G
NF-G
Quantile


Home
About
Features
Compounds
Partners
Forum
Contact
Log in
Standard
Compare
Structures
Views
Standard
Compare
Structures
Libraries
Reference Library
Search
Spectrum
Tree
Structure
Monoisotopic Mass
Peak
Precursor
Name
Search Results
Reference Library
Filter
mannitol
Results for 'mannitol'
No: 1501
D-Mannitol 1-phosphate
Monoiso. Mass: 262.04537
Thermo
No: 1527
D-(-)-Mannitol
Monoiso. Mass: 182.07904
Thermo
Spectral Tree
Filtered
Recalibrated
1/2 MS1 Combined Scans #31, 89
Structure
$C_6H_{14}O_6$
Filtered Spectrum
[M - H]⁻ 181.07204
[M + FA - H]⁻ 227.07729
129.05592
311.1349
Precursor Structure
To see precursor structure please select tandem spectrum in Spectral Tree
Blue Structure: Heuristic Prediction
Brown Structure: Quantum Chemical Prediction
Filtered Spectrum
85.02987
[M-H]⁻
m/z 181.07207
CID 35: IW 3
m/z 163.06149
CID 40: IW 3
m/z 113.02473
CID 40: IW 2

ELECTROPHORESIS

A

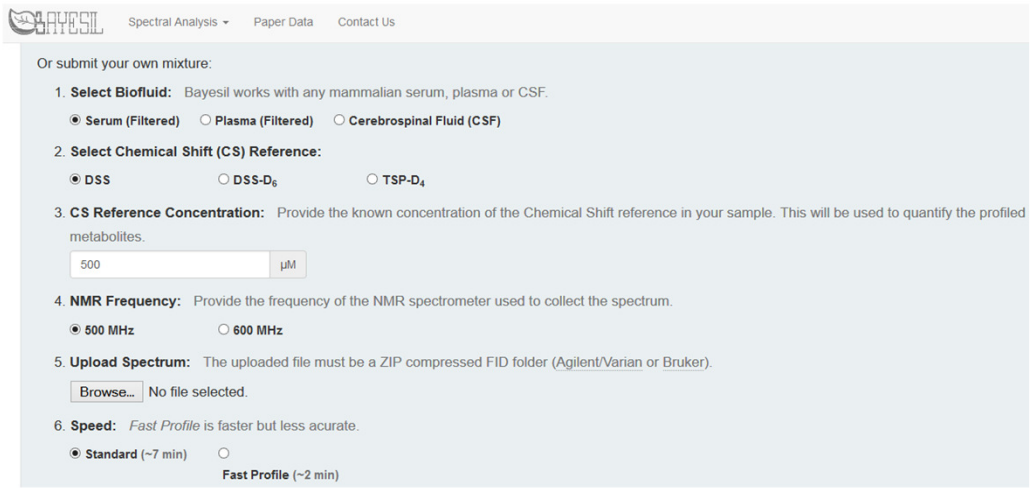


B

Date: 2015-08-04 23:29:53 UTC
FID: Example_3_Biological_Serum_Bruker_600MHz.fid.zip
Biofluid: Serum (Filtered)
NMR Frequency: 600 MHz
DSS Concentration: 500.0 ÂµM
Fast Profile: No

HMDB ID	Compound Name	Concentration (ÂµM)	Threshold	Confidence Score
-	DSS	500	0.6	
HMDB00001	1-Methylhistidine	4.4	5.7	5
HMDB00008	2-Hydroxybutyric acid	35.2	5.7	9
HMDB00042	Acetic acid	24.8	4	10
HMDB00043	Betaine	42.4	5.7	10
HMDB00060	Acetoacetate	4.3	5.7	8
HMDB00062	Carnitine	28.8	5.7	9
HMDB00064	Creatine	38.1	5.7	7
HMDB00094	Citric acid	192.9	5.7	10
HMDB00097	Choline	36.3	1.7	10
HMDB00108	Ethanol	35.2	5.7	9
HMDB00122	D-Glucose	1788.2	28.7	10
HMDB00123	Glycine	406.5	5.7	10
HMDB00131	Glycerol	96.9	5.7	9
HMDB00142	Formate	26.4	4.6	9
HMDB00148	L-Glutamic acid	237.2	5.7	10
HMDB00157	Hypoxanthine	68.3	5.7	10
HMDB00158	Tyrosine	81.1	5.7	10
HMDB00159	L-Phenylalanine	82.9	5.7	10
HMDB00161	L-Alanine	310.9	11.5	10

C



RIKEN PRIMe provides web-based data analysis and visualization tools for public access. Users may analyze datasets from both AtMetExpress and MeKO, as well as upload custom datasets.

Analysis parameters

Data source: *MeKO*

Dataset: *All*
Normalization: *CRMN*

Downloads

Limma table TSV

Edit parameters

Summary (MVA) | Differential Accumulated Metabolites (UVA)

Hierarchical Cluster | Heat Map | Principal Component Analysis

hclust Method

Complete

dist Method

Manhattan

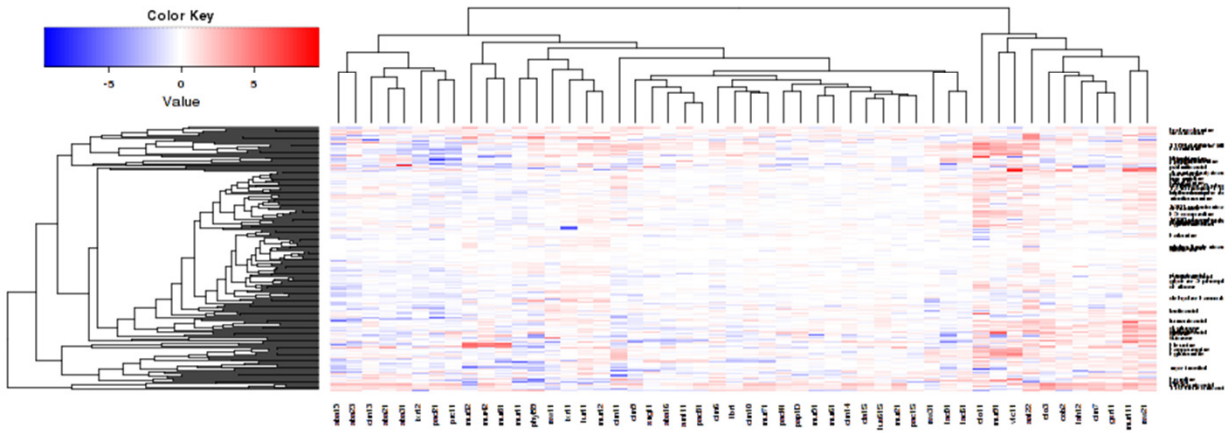


Table 1. Table enlisting the metabolomics tools and resources developed during the period 2014-15, consisting of the name, and displaying their platform dependencies in terms of analytical input and computational dependencies, web address (URL), and their reference if available. Abbreviations used: Any: input data from different kinds of analytical platforms, GC: gas chromatography, Internet: tool running on server and assessable by online access, LC: liquid chromatography, MS: mass spectrometry, MS/MS: mass spectrometry fragmentation, NMR: nuclear magnetic resonance spectroscopy, R: R package, UV-DAD: ultraviolet diode array detection, - : not applicable/available.

Name	Platform Dependencies		URL	Reference
	Analytical Input	Computational		
Data handling and preprocessing tools				
MetMSLine	LC-MS	R	http://wmbedmands.github.io/MetMSLine/	17
MRM-DIFF	LC-MS	Windows	http://prime.psc.riken.jp/	20
MRMPROBS	LC-MS	Windows	http://prime.psc.riken.jp/	21
FragPred	MS	R	http://pattilab.wustl.edu/software/FragPred/index.php	23
MUSCLE	LC-MS	C++	http://www.muscleproject.org/	24
IsoMS	LC-MS	R, Windows	www.mycompoundid.org/IsoMS	25
MyCompoundID.org	MS	Internet	http://mycompoundid.org/	26
IsoMS Quant	LC-MS	R, Windows	www.mycompoundid.org/IsoMS	27
isoMETLIN	MS/MS	Internet	http://isometlin.scripps.edu/	28
IPO	LC-MS	R	https://github.com/glibiseller/IPO	30
Massifquant/ Optimize-it	MS	Bioconductor	https://github.com/topherconley/optimize-it	31
intCor	LC-MS	R	http://b2slab.upc.edu/software-and-downloads/intensity-drift-correction/	37
Peak-group-alignment	LC-MS	R	https://github.com/joewandy/peak-grouping-alignment	38
Parametric time warping	LC-MS, LC-UV-DAD	R	https://github.com/rwehrens/ptw	42
PeakANOVA	LC-MS	R	http://research.ics.aalto.fi/mi/software/peakANOVA/	43
MSPrep	MS	R	http://sourceforge.net/projects/msprep/	53
Statistical tools				
DeviumWeb	Any	R, Internet	https://github.com/dgrapov/DeviumWeb	47
EigenMS	MS	R, Matlab	http://sourceforge.net/projects/eigenms/	50
RepExplore	Any	Internet	http://www.repexplore.tk	51
Normalyzer	Any	Internet	http://quantitativeproteomics.org/normalyzer	52
BioStatFlow	Any	Internet	http://biostatflow.org/	-
Annotation tools				
POCHEMON	LC-MS	Matlab	http://www.ru.nl/science/analyticalchemistry/research/software/	54
BioSM	MS/MS	Java	http://metabolomics.pharm.uconn.edu	61
BioSMXpress	-	-	http://engr.uconn.edu/~rajasek/BioSMXpress.zip	62
CFM-ID	ESI-MS/MS	Internet	http://cfmid.wishartlab.com/	64
GS-align	-	C++	http://www.glycanstructure.org/gsalign	66
Cosmiq	LC-MS, GC-MS	R	http://www.bioconductor.org/packages/release/bioc/html/ cosmiq.html	67
Feature Credential	MS/MS	R	http://pattilab.wustl.edu/software/credential/	68
HAMMER	MS/MS	Java, Python	http://www.biosciences-labs.bham.ac.uk/viant/hammer/)	70
HR3	LC-MS, GC-MS	Windows	www.metalign.nl	71
LipidPro	MS/MS	Windows	http://www.neurogenetics. biozentrum.uni-wuerzburg.de/en/project/services/lipidpro/	72
MAGMa	MS/MS	Java	https://www.emetabolomics.org/	74
MAIT	MS	R	http://b2slab.upc.edu/software-and-downloads/metabolite-automatic-identification-toolkit/	76
Metabolome searcher	-	Internet	http://procyc.westcent.usu.edu/cgi-bin/MetaboSearcher.cgi	77

MET-COFEA	LC-MS	Windows	http://bioinfo.noble.org/ manuscript-support/met-cofea/	78
MetAssign	LC-MS	Java	http://mzmatch.sourceforge.net/MetAssign.php	79
MS2Analyzer	MS/MS	Java	http://fiehnlab.ucdavis.edu/projects/MS2Analyzer/	80
MIDAS	MS/MS	Internet	http://midas.omicsbio.org	81
mzCloud	MS/MS	Internet	https://www.mzcloud.org/	82
MS-DIAL	LC-MS	Windows	http://prime.psc.riken.jp/	88
mzGroupAnalyzer	LC-MS	Matlab	http://www.univie.ac.at/mosys/software.html	90
ProbMetab	LC-MS	R	http://labpib.fmrp.usp.br/methods/probmetab/	91
RAMClustR	MS/MS	R	https://github.com/cbroeckl/RAMClustR	92
Pathway and networks analysis, and biological interpretation tools				
PathCaseMAW	Any	Internet	http://nashua.case.edu/PathwaysMAW/Web	55
MINE	-	Java, Perl, Python	http://minedatabase.mcs.anl.gov/#/home	83
PathWhiz	-	Internet	http://smpdb.ca/pathwhiz	95
TrackSM	MS/MS	Matlab	http://metabolomics.pharm.uconn.edu/?q=Software.html	96
MarVis-Suite	Any	Java	http://marvis.gobics.de	97
InCroMAP	Any	Java	http://www.cogsys.cs.uni-tuebingen.de/software/InCroMAP	98
iPEAP	Any	Windows, Java, R	http://www.tongji.edu.cn/~qiliu/ipeap.html	99
kpath	Any	Internet	http://browser.kpath.khaos.uma.es/	100
Pathomx	Any	Python, Matlab, R	http://pathomx.org/	101
MetaMapR	Any	R, Internet	http://dgrapov.github.io/MetaMapR/	105
Metabnet	LC-MS	R	https://sourceforge.net/projects/metabnet/	111
MetaNET	Any	Galaxy	http://metanet.osdd.net	121
Funrich	-	Internet	http://www.funrich.org/	122
Network portal	Any	Internet	http://networks.systemsbiology.net	123
SimIndex	Any	Python	http://tyolab.northwestern.edu/tools/	125
BiNChE	-,,	Internet	http://www.ebi.ac.uk/chebi/tools/binche/	126
PlantSEED	-,,	Internet	http://bioseed.mcs.anl.gov/~seaver/FIG/seedviewer.cgi?page=PlantSEED	127
SPICA	LC-MS	-	http://cmcr.columbia.edu/metabolomics/informaticstools.html	128
KiMoSys	Any	Internet	http://kimosys.or	129
Integrated Interactome System	Any	Internet	http://www.lge.ibi.unicamp.br/lnbio/IIS/	130
MetaDB	LC-MS, GC-MS	R	https://github.com/rmylonas/MetaDB	164
GC-MS-based tools				
MetaMS	GC-MS	R	http://www.bioconductor.org/packages/release/bioc/html/metaMS.html	163
Maui-VIA	GC-MS	Java	http://bimsbstatic.mdc-berlin.de/kempa/software/kempaSoftware.html	181
PScore	GC-MS	R	http://raphaelaggio.github.io/	184
BiPACE 2D	GC-MS, 2D GC-MS	Java	http://maltcms.sf.net/	185
NMR-based Tools				
Focus	NMR	MATLAB	http://www.urr.cat/FOCUS	187
Metabnorm	NMR	R	http://sourceforge.net/projects/metabnorm/	189
BATMAN	NMR	R	http://batman.r-forge.r-project.org/	188
Bayesil	NMR	Internet	http://bayesil.ca/	190

SENECA	NMR	Java	http://sourceforge.net/projects/seneca/	191
COLMAR	NMR	Internet	http://spin.ccic.ohio-state.edu/index.php/colmar	192
$^{1}H(^{13}C)$-TOCCATA	NMR	Internet	http://spin.ccic.ohio-state.edu/index.php/toccata2/index	193
mQTL.NMR	NMR	R	http://www.ican-institute.org/tools/	194
MVAPACK	NMR	Matlab	http://bionmr.unl.edu/mvapack.php	195
ChemoSpec	NMR	R	http://cran.r-project.org/web/packages/ChemoSpec/	196
Library, databases and others				
TAPIR	MS/MS	Python	https://github.com/msproteomicstools/msproteomicstools	131
TMDB	-	Internet	http://pcsb.ahau.edu.cn:8080/TCDB/	138
BioPhytMol database	-	Internet	http://ab-openlab.csir.res.in/biophytmol/	139
EssOilDB	-	Internet	http://nipgr.res.in/Essoildb/	140
mVOCs	-	Internet	http://bioinformatics.charite.de/mvoc	141
domdb	-	SQL	https://github.com/joefutrelle/domdb	143
PhenoMeter	-	Internet	https://www.metabolome-express.org/phenometer.php	144
T3DB	-	Internet	www.t3db.ca	145
SwissLipids	MS	Internet	http://www.swisslipids.org/	146
Metabolonote	-	Internet	http://metabolonote.kazusa.or.jp/	161
KOMICS	-	Internet	http://www.kazusa.or.jp/komics/	162
QTREDS	-	Ruby, Rails	http://qtreds.crs4.it	165
MASTR-MS	-	Internet	https://mastr-ms.readthedocs.org/en/latest/	167
Yabi	-	Internet	http://ccg.murdoch.edu.au/yabi/	169
jmzTab	-	-	http://mztab.googlecode.com	172
PolySearch2	-	Internet	http://polysearch.ca	175
SpeckTackle	MS/MS, NMR	Java	https://bitbucket.org/sbeisken/specktackle	176
BioMet Toolbox 2.0	-	Internet	http://biomet-toolbox.org/	177
Metabolite Imager	-	Internet	www.metaboliteimager.com	179
EXIMS	MALDI-MS	Matlab	http://exims.sourceforge.net/	180
Multifunctional tools				
XCMS Online	LC-MS, GC-MS	Internet	https://xcmsonline.scripps.edu/	201
Mass++	LC-MS, GC-MS	Java	http://www.first-ms3d.jp/english/	203
MASSyPup	LC-MS, GC-MS	Linux	http://www.bioprocess.org/massypup	204
MetaboNexus	Any	Windows, Java	http://www.sph.nus.edu.sg/index.php/research-services/research-centres/ceohr/metabonexus	205
MetaboLyzer	LC-MS	Python	https://sites.google.com/a/georgetown.edu/fornace-lab-informatics/home/metabolyzer	206
Workflow4Metabolomics	Any	Galaxy	http://workflow4metabolomics.org	207
Haystack	LC-MS	Internet	http://binf-app.host.ualr.edu/haystack/	208
ALLocator	LC-MS	Internet	https://allocator.cebitec.uni-bielefeld.de	209
MeKO	GC-MS	Internet	http://prime.psc.riken.jp/meko/	210
MassCascade	LC-MSn	KNIME	https://bitbucket.org/sbeisken/masscascadeknime/wiki/Home	211